Anders Klevmarken

Statistical Methods for the Analysis of Earnings Data

THE INDUSTRIAL INSTITUTE FOR ECONOMIC AND SOCIAL RESEARCH

The Industrial Institute for Economic and Social Research

is an independent non-profit research institution, founded in 1939 by The Swedish Employers' Confederation and The Federation of Swedish Industries.

Objectives

To carry on research into economic and social conditions of importance for industrial development in Sweden.

Activities

The greater part of the Institute's work is devoted to long-term problems, especially to long-term changes in the structure of the Swedish economy particularly within manufacturing industry. This also includes continuous studies of the development of private consumption in Sweden and projections of demand for various products. Research results are published in the series issued by the Institute.

Along with the long-term research work the Institute carries out investigations concerning special problems and performs certain services to industrial enterprises, organizations, governmental agencies, etc.

Address

Industriens Utredningsinstitut
Storgatan 19, Stockholm, Box 5037,
S-102 41 Stockholm 5, Sweden
Tel. 08/63 50 20

Statistical Methods for the Analysis of Earnings Data

The Industrial Institute for Economic and Social Research

Statistical Methods for the Analysis of Earnings Data

with special application to salaries
in Swedish industry

Anders Klevmarken

Almqvist & Wiksell, Stockholm, Sweden

ISBN 91-7204-004-1

Printed in Sweden by Almqvist & Wiksells Boktryckeri AB, Uppsala 1972

CONTENTS

FIGURES

TABLES

FOREWORD

Statistical studies of earnings are important for many purposes. In the complicated and technical process of present day negotiations comparisons of wages and salaries are done by more and more refined methods. Similar comparisons are central both in professional and political debate about income distribution and inequality in standard of living. Studies in the structure of earnings are also important for the analysis of consumptions and savings behavior and for educational planning.

The complex interaction between all those factors which determine the structure of earnings makes the analysis a difficult task. In empirical work it is thus urgent to try various methods of general applicability which catch significant features of the earnings structure. The present study is a statistical analysis of the salary structure in Swedish industry. Although the methodological aspects of the study are important, the empirical illustrations of the methods of analysis give results which carry over as suggestions for future research.

The study is made by fil.lic. Anders Klevmarken. It started while the author was employed by the University of Stockholm and the National Central Bureau of Statistics. The study was then completed at the Industrial Institute for Economic and Social Research (IUI) and jointly sponsored by all three institutions. The statistical data have been obtained from the salary statistics of the Swedish Employers' Confederation (SAF) and the Swedish Association of Graduate Engineers (CF). We are grateful to these two organizations, without the generous support of which this study would have been almost impossible to complete. We also wish to express our appreciation to the seminar at the Institute of Statistics, University of Stockholm and the seminar at The Stockholm School of Economics for their comments and suggestions. The study has also benefitted from financial support from The Swedish Council for Social Science Research.

Stockholm in April 1972

Sten Malmquist	Lars Nabseth	Ingvar Ohlsson
Institute of Statistics	The Industrial Institute	National Central
University of Stockholm	for Economic and Social	Bureau of Statistics
	Research	

PREFACE

The research underlying this monograph started some years ago, initiated by Sten Malmquist, head of the Institute of Statistics, University of Stockholm, and Carl Johan Åberg, former head of the Forecasting Institute, National Central Bureau of Statistics. As the study advanced preliminary results were documented in various working papers and research reports. Chapter 3 on labour composition and mobility contains results obtained already in 1967. The analysis of age-earnings profiles in Chapter 4 received its present form after several revisions of a preliminary draft from 1968 and one version was read at the European Meeting with the Econometric Society in Barcelona in 1971. Many of the results from the cross sectional analysis of salary differences in Chapter 5 were presented in a research report in 1968. This monograph thus contains selected results from previous reports now put together and extended with new ones.

It has indeed been a great privilege to work in the stimulating environment of three research institutes: the Institute of Statistics, the Forecasting Institute and the Industrial Institute for Economic and Social Research (IUI). My greatest obligation is to Sten Malmquist for all his encouragement and patient guidance. His generous interest and our frequent discussions have been a great support, in particular when doubts about my progress were persistent. I am also indebted to Thora Nilsson, head of the Forecasting institute, and Lars Nabseth, director for the Industrial Institute for Economic and Social Research, who have both actively contributed to my work, not only by »administrative» support but also by comments and suggestions to improvements.

Harry Lütjohann and Staffan Lundquist were exceptionally generous struggling through unorganized early versions of the manuscript, discussing substantive and expository issues at length, and making a great number of invaluable contributions. No doubt, the book improved very much as a result of their helpful efforts.

I have been unusually fortunate in receiving many useful suggestions from Tore Dalenius, Gunnar Eklund, Gunnar Eliasson, Karl-Olof Faxén, Siv Gustafsson, Jan Henriksson, Erland Hofsten, John Quigley, Erik Ruist and Carl Johan Åberg.

I also wish to express my special appreciation to Hans Löfgren and his collegues at The Swedish Employers' Confederation (SAF) who never refused assistance in providing and interpreting SAF's statistical data, nor in commenting on preliminary drafts of the manuscript. Thanks are also due to Jon-Erik Eriksson and his collegues at The Swedish-Association of Graduate Engineers (CF) who helped with data and comments on empirical results. I am grateful to Kurt-Allan Wallgren for his assistance in comparing the SAF-data and the CF-data (Chapter 2) and to Sune Robin for providing data from the job-indicator in Dagens Nyheter. I also wish to thank Bengt Cedheim who made a great job in programming all the computations for Chapter 5, and L. Gruber who improved my English.

Grateful thanks also go to editorial and secretarial help at all three institutes. In particular I wish to mention the generous assistance of Ruth Ellerstad and Wera Nyrén at IUI in preparing the manuscript for the printer.

I am of course responsible for any remaining errors of fact or interpretation.

Stockholm in April 1972
Anders Klevmarken

CHAPTER 1

GENERAL INTRODUCTION AND SUMMARY

Wage and salary differences are of general concern and in the debate the »facts» presented to support arguments and interpretations of statistics are often contradictory, or at least they seem to be. In the authors's opinion this is not usually a result of a conscious attempt to misuse statistics for the sake of a good cause, but rather the effect of different definitions and different methods of analysis. Lack of data, lack of good methods and insufficient understanding of the assumptions behind existing methods leave a wide range for »expert» interpretations.

In their textbook Wallis and Roberts (Wallis and Roberts [1956]) give many simple but illustrative examples. One of them is example 77C (p. 77).

»A manufacturing plant found that the average monthly earnings of its employees had fallen 8 percent during a certain period. This might seem to »prove» that earnings had gone down. As a matter of fact, however, the earnings of every single employee were exactly 10 percent higher than at the beginning of the period. The reason the average earnings fell despite this increase was that many of the higher-paid employees were dropped at the time the increase was made, so that the new average included only lower-paid workers.»

Another example is obtained from the negotiations and conflict between the Swedish Confederation of Professional Associations (SACO), the National Federation of Government Officers (SR) and the Swedish National Collective Bargaining Office during the winter and spring of 1971. In this tense situation contradictory »facts» showing salary increases during past periods of varying length and in comparison to other groups in the labour market were presented by both sides. In negotiations comparisons of wages and salaries between groups of employees have traditionally been made, whenever possible, after standardization for differences in age and kind of job. The basic principle underlying the interest in standardization for differences in the kind of job is that of equal pay for equal job. This principle does not say anything about the wage or salary difference between two different jobs and it does not necessarily conflict with a policy aiming towards an equalization of wages and salaries. Such a policy is nothing new in negotiations, but in the negotiations during the winter and spring of 1971 these aspects were spotlighted and given a new emphasis. While SACO and SR, which represent civil servants and teachers at middle and high levels in the public sector, presented figures to show what small nominal increases and even real decreases their member groups had received during the duration of the previous agreement, new statistics were suddenly thrown into the debate showing that these groups had in general obtained increases exceeding that of any other group. From the statistical point of view it is interesting to note that they were obtained by a method of comparison rarely used before in negotiations in Sweden. These new data were obtained from a public investigation (Eriksson [1970]) where each individual in a sample of SACO and SR members was followed over a number of years and their salaries recorded. The average salary increase obtained from these data was high because promotion from one job level to a higher one in the career of a civil servant was included in the measurement. There was thus no standardization for differences in job level in this comparison.

There is no easy answer to what is the proper comparison, but it is important that the nature of different methods of measurement and the differences between these should be clearly understood which was probably not the case during the conflict. (See also Klevmarken [1971].)

1.1 AIMS

The two examples above are both related to negotiations, but the main purpose of this study is more general than merely that of criticizing statistical practice in this particular field of application; it is rather to suggest some methods for the analysis of income and earnings statistics of a wider applicability and to illustrate empirically how these methods work. This study is not intended to solve *one* particular problem by *one* method especially designed for this problem. On the contrary, the methods suggested in this study are believed to be useful in several different applications, and some examples will be given.

The methodological aspects have offered more than enough problems of interest for one study. Analysis of data from a subject matter point of view is therefore not a primary objective of this study. The empirical analysis first of all serves as an illustration of the methods. However, this does not mean that the empirical results obtained are not interesting in their own right. Although it is the methodological aspects of the study which are stressed, its purpose is in this sense twofold. This double objective is more a problem of presentation than a real problem, because a discussion of methods must always have some relation to applications. Several examples of possible applications as well as references to work carried out by others are given below. Most of them are taken from the fields of economics and salary negotiations, but there are also some references to other fields of application.

The construction of the models used in this study is naturally influenced by subject matter studies performed by others. This is for instance true for the choice of variables which is also limited by available data. As the statistics used were originally collected for use in negotiations, we have some guarantee that the variables are at least relevant for this purpose. The present applications of the methods suggested may be called descriptive in the sense that little of economic or institutional explanation is offered of the particular features found in the data in addition to what is now inherent in the models. Although they are not based on a *particular* economic or industrial relations theory, the models can easily be modified and expanded to give answers to economic problems. They can thus serve as a basis for an economic analysis. Some indications in this direction are given in particular in chapter 4, but it is beyond the scope of this study to pursue the analysis in that direction.

1.2 A READER'S GUIDE

Except for chapters 1 and 2 each chapter is organized in one theoretical and one empirical part. In order that the reader may benefit from the text in an empirical part it is not necessary that he should read the corresponding theoretical part in detail, but it is assumed that the reader is acquainted with the notation and main characteristics of the model. It is not necessary to read all chapters in one sequence since each chapter is relatively selfcontained.

Because of the twofold purpose of the study and the possible difference in interest of the readers it may be appropriate to give a relatively detailed guide to the following chapters and summarize the results chapter by chapter.

1.2.1 *A reader's guide to chapter 2*

Even if the main objective is methodological it is necessary to know the limitations and nature of the statistical data in order that appropriate methods may be designed. It needs hardly be mentioned that this knowledge is a prerequisite for interpretation of the empirical results. The data used are salary statistics from the Swedish Employers' Confederation (SAF) and from the Swedish Association of Graduate Engineers (CF) which collect salary data for their negotiations. For readers not familiar with Swedish negotiations for salary earners, the very brief survey of institutions and negotiation practice at the beginning of chapter 2 may facilitate understanding of the interest in a study of the salary structure and also the instrumental rule of the statistics. The remainder of chapter 2 is a rather detailed survey of statistical data, its collection, definitions and limitations. As the SAF data only cover salary earners employed by members of SAF there are groups of employees which enter and leave the records. It is important that readers who want to interpret the empirical results, in particular those in chapter 3, should read chapter 2. Similarly, the statistics do not cover employees at top management level which gives a selection bias in the estimates of salary increases in chapter 4 and of salary differences in chapter 5. This data problem is also treated in chapter 2.

1.2.2 *A reader's guide to chapter 3*

Chapter 3 serves as a general background to the empirical analysis in chapters 4 and 5. The reader who wants to interpret the empirical results will find useful information on for instance age, education and job composition. The distribution of the number of employees over the educational qualifications, jobs, age intervals and industries presented shows what educational qualifications, jobs and so forth are most frequent and for this reason most worthwhile to analyse in the following chapters. Chapter 3 also contains some information on mobility among salary earners which may be interesting as a background in particular to the investigation of changes in the salary structure in chapter 5.

For practical reasons composition and changes in composition with respect to age, education, job and so forth are analysed for two variables at a time, usually education and another variable. The description follows a scheme now exemplified by its application to the variables education and industry. First the educational distribution (composition) is given for all industries together for two or three years. The change in composition between these years is also presented in detail for all industries together. The educational distribution for each industry separately is only presented for the last year. Changes in composition for a single industry are measured by two descriptive statistics. The industries which according to these statistics show a change in composition that deviates most from the change registered for all industries are singled out for a closer examination of the nature of the observed change, while the compositional change in the remaining industries is considered to be approximately equal to the overall change.

The two descriptive statistics used have no particular relation to the analysis of the salary structure in chapters 4 and 5. One objection to their introduction then is that their fitness for use in the description of compositional changes in the population of salary earners is difficult to evaluate, but on the other hand, their simplicity and affinity both with the usual measures of central tendency and variation and with index numbers may justify their introduction and application. As these statistics do not necessarily have to be applied to earnings data there is a wider applicability of this approach.

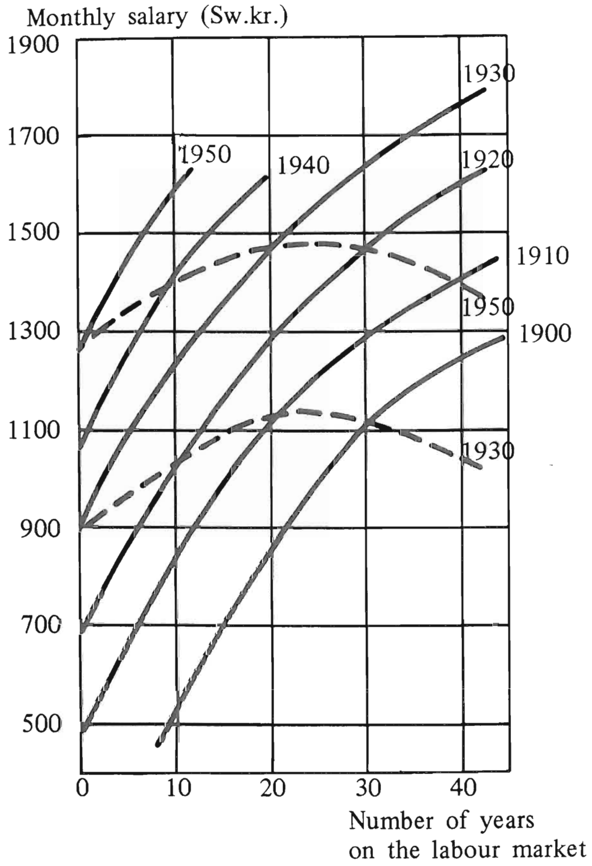
In chapter 3 it is shown that the average annual turnover of salaried employees is 15–20 % of all employed salary earners. This individual mobility and an increased supply from schools, universities and other sectors, increased the total number of salary earners with academic or vocational high school training employed by SAF members by more than 60 % over the period 1957–1968. Although the aggregate distributions by age, kinds of education, job and industry only show changes in frequency of not more than a few percentage units there are some decided changes in composition. The general increase in the number of employees is unevenly distributed among industries. In the period 1957–1968 this more than doubled in Building and construction, while it decreased in Beverage and tobacco and in Textile industry. The number of employees with a nonacademic education increased more than those with an academic degree. The age distribution also changed. The number of employees below 30 years and between 45 and 60 years increased more than other age groups. Yet another change in composition is the decrease in the proportion of employees at high job levels. Although no analysis of the causes underlying mobility and changes in salary structure is performed in this study the reader may find this information relevant for an explanation of the rigid salary structure found in chapter 5.

1.2.3 *A reader's guide to chapter 4*

In econometric literature a distinction is made between cross section analysis and time series analysis. Nevertheless differences between individuals measured in cross sections are commonly interpreted as if the same differences had been observed from a time series. For instance, the income elasticity of consumption estimated from a household survey is sometimes used as a measure of the effect on consumption of an increase in income from one point of time to another. Such an agreement of results from cross sections and time series can not of course be expected to hold in general (Malmquist [1948],pp.62 ff, Klevmarken [1970]).

Since there is an interest in comparisons of earnings between points of time and between individuals, it was natural to develop a model which distinguishes between these two aspects but at the same time links them together. The result is the model developed in chapter 4. The simple observations behind it can most easily be explained by figure 1:1. Suppose a group of individuals is defined by some common characteristics and that it is divided into cohorts

Figure 1:1. *Cohort and cross sections profiles*



according to when they started work. Assume also that those who started work in 1910 received the average initial monthly salary of Sw.kr. 500. As they grow older their average salary increases and in 1950, when they have worked for 40 years, they reach the salary of Sw.kr. 1400. Since starting salaries increase from one year to another, those who start work later accumulate their increases from a higher initial level. For instance, those who started in 1920 received a starting salary of Sw.kr. 700 and those who started in 1950 Sw.kr. 1300. If each solid curve is followed to a point which corresponds to a particular calendar year, for instance 1950, and all these points are joined, the broken curve marked 1950 is obtained. This curve describes differences between (groups of) individuals at one and the same point of time (cross section profile), while a solid curve describes the salary path of a group through time (cohort profile). As is shown by the figure the two kinds of curves (profiles) can be very different.

Age-earnings profiles are of interest in several applications. They are central in the human capital approach in economics, where the profiles are looked upon as a result of investment in human capital (Becker [1962], Mincer [1970]),and they are used in educational planning to calculate rates of return for education (Blaug, Peston & Ziderman [1967], Siegfrid [1971]).★ They are also used in consumption and savings theory to explain the life-cycle pattern of consumption and savings (Lydall [1955], Tobin [1967], Thurow [1969], Klevmarken [1970]). The profiles are of importance in collective bargaining. Labour unions organizing employees of higher education (academic and professional education) have in their negotiations with employers used the characteristics of the age income profile as an argument for high salaries compared to those groups which do not have a high education and thus enter the labour market earlier. Sometimes the comparison between the groups is done on the basis of lifetime incomes (SACO [1968]). The usual practice of making these calculations from cross section profiles has been criticized in for instance Miller [1965] and Ben-Porath [1966]. The discussion about income inequality in Sweden and the example from the negotiations in 1970/71, which was given above is another example of the application of age-earnings profiles. Yet another application is described in chapter 2, namely that cross section profiles are used in the negotiation process in Swedish industry to determine the so called statistical salary increases due to age (Lind [1963], SAF [yearly]).

In his work on age-earnings profiles Fase (Fase [1969]) mentions actuarial science as another field of application. The profiles may then be used to estimate earnings lost for instance because of injury in an accident. (References are given in Fase [1969].)

Cohort and cross section profiles with the same basic characteristics as age-earnings profiles also have applications in other sciences. That cross section data should be supplemented by cohort data in order to avoid misleading conclusions has long been recognized in the science of demography. As an early example reference may be made to the report of the 1860 Swedish population census, signed by F. Berg, the head of the then recently organized Swedish Central Office of Statistics. For a long time, however, the dominance of the life table which in its most current form is nearly always based on cross section data has had the result that the cohort approach has been neglected

★ For a criticism see for instance Merrett [1966].

to some extent. Modern development started with Whelpton [1954] who pointed out that the cohort approach is fundamental for studying fertility and who was probably also one of the first scientists to introduce the term cohort in this context. A more recent Swedish reference is Hofsten [1971].*

Yet another example of a combined cohort and cross section analysis is found in a recent article about the age distribution of cancer (Doll [1971]). In a diagram similar to figure 1:1 the author demonstrates that the annual death rates in lung cancer in England and Wales show the reversed U-shaped cross section profile by age so well known from economic applications. This is so because the death rate increases not only with age but also for each successive cohort.

In applications the data available for derivation of profiles are very often insufficient and inaccurate. Usually only one or a few cross sections are available and there may be changes in the cross section profiles because some observational units are reported one year but not another. Data on individuals may be very expensive or impossible to obtain while avarages, for instance average incomes or average earnings, may be more readily available. A model which is specified in such a way that estimation is possible from this type of data and still permits projections of complete profiles would be very useful in all applications mentioned.** The aim of chapter 4 is therefore to develop a statistical framework for analysis and projection of age-earnings profiles with a limited supply of data.

The chapter starts with a formal representation of an individual earnings path which can be applied to any data. The individual path is seen in this representation as a deviation from the average path of a well defined group. The average path is built up of an average initial salary to which average yearly salary increases are accumulated, Initial salaries also change from one year to another. A reason for the immediate introduction of an average path is that the data to be used are average salaries.

In the later sections of chapter 4 certain simple but rather natural restrictions are imposed on the formal representation which give a few alternative versions of the same basic model. In the first model the salary profiles are

---

* I am indebted to E. Hofsten for comments and references on the application of cohort and cross section analysis in demography.

** I am indebted to T. Dalenius who suggested that such a model could be used for analytic evaluation of data. For instance, cross section data different from the data used for estimation could be confronted with the predictions from the model to discover errors in the statistics.

analysed as a function of active age, i.e. number of years in the labour market, and this model serves as a kind of prototype model. In the second model the salary increases are a function of both active age and physical age. The third model which is a special case of the second, describes earnings as a function of physical age only. Before the empirical application to pooled cross section and time series data from CF and SAF is presented, there is a section about individual variability and the stochastic properties of the models. It is for instance shown that the individual dispersion around the average path, as measured by a standard error, is of the same relative magnitude as the average annual increases.

The application of the model to data from CF shows that active age is more important for the determination of salary increases than physical age. This is a result which is consistent with the human capital theory. When the physical age profile model is applied to SAF data grouped by education, the results show that the estimated average percentage shifts in the cross sections are almost the same for all educational groups. Disregarding a possible bias in the estimates this result can also be interpreted as an almost equal increase in initial salary for the different educational qualifications. The estimates of the age dependent salary increases show more dissimilarities between educational groups. The profiles have a more pronounced peak for employees with a university degree than for those without. To some extent this may be an effect of the selection bias previously mentioned.

Under the assumptions of the model used, it is also shown that a lifetime salary calculated from a cross section profile with a certain discount factor, numerically equals the lifetime salary obtained from a cohort profile discounted by a factor which is the sum of the average percentage increase in initial salary and the discount factor used in the cross section calculations. The estimated models are also used for numerical calculations of lifetime salaries as well as comparative rates of return for education before tax. As the assumptions used in these calculations are restrictive they primarily serve as illustrations. The results may be used as a characterization of the sample period but preferably not as predictions. The restrictive assumptions which now make predictions doubtful can, however, be modified and will make the models more realistic. The parameters of the models, which are interpreted as salary increases, are now assumed to be invariant to factors associated with calendar time but independent of cohort. As the systematic pattern of the residuals analysed in section 4.3.3 shows, this is a rather restrictive assumption. One way of obtaining the desired modification would be to expand the present models with an explanation of the inter-yearly variations in salary increases in terms of supply and demand for labour, consumer price increases and other factors. By this kind of expanded model it might be possible to make meaningful predictions of changes in the shape of the profiles and then also of future lifetime salaries.

1.2.4 *A reader's guide to chapter 5*

Although it has not been possible within the present study to incorporate explicitly economic growth factors and price increases in the models and in this way to investigate whether, in the terminology of Ben-Porath [1966], economic growth and inflation are neutral with respect to the shape of cross section profiles, the salary structures of isolated cross sections are analysed and compared in chapter 5. The definition of structure is mainly determined by the limited set of variables in the salary data, which of course reflects the concepts and comparisons of interest in negotiations. In chapter 4 we only consider salary differences due to education and age, but in chapter 5 differences due to job level, job family, industry and cost of living area are analysed in addition. It is quite natural to see promotion from one job level to another as one component of the age-related salary increases in chapter 4 while the other factors are mainly used for stratification in the same way as education.

Chapter 5 is, however, not only an investigation of the stability of the salary structure as observed in cross sections, but it is also intended as a general purpose description of the salary structure itself in the sense that it should be possible to use the description for many different comparisons of salary differences, possibly by different users. A descriptive problem usually involves a conflict between ease of presentation and loss of information. The approach taken here is firstly to give an overall survey of the salary structure by using a simple and rather restrictive model and secondly to go into some detail which for practical reasons has to be done on subsets of data.

The statistical model used is the general linear model. Application of this model makes comparisons possible in a general way. As comparisons between earnings of different groups of employees are important in negotiations, this is a field in which application of the methods in chapter 5 could be very useful. The brief survey of negotiation practice in section 2.2 will help to clarify this point. Another field of application is the analysis of income distributions in which the general linear model can help to throw light on the role of important socio-economic variables in shaping the distribution of income from employment as shown in Hill [1959].

Chapter 5 starts with a presentation of the general linear model of less than full rank by an example with two factors. It continues with an analysis of some specific properties of this model. Since it is not a full rank model the parameter values are not all unique and estimable. There are, however, linear combinations of parameters which are unique and estimable. Traditionally the general linear model of less than full rank has been used mainly to test an estimate estimable linear combinations, e.g. differences between two parameters. In view of the general purpose of this study, focusing on estimable linear combinations is not appropriate. We aim at a general presentation of the salary structure which makes many comparisons feasible. The approach chosen is to impose constraints on the parameters in order to obtain unique parameter values, which may be seen as a standardization of the parameters. The constraints have to be chosen according to certain rules, but there are still an infinite number to choose from, i.e. the parameters can be standardized in many different ways. To investigate estimable functions any of these constraints can be used to obtain a solution to the normal equations, but when an estimate of each parameter is to be presented one standardization may be more convenient than another. The choice of constraints within the class of feasible constraints determines the interpretation of the parameters and the corresponding estimates. In a particular application, one interpretation may be preferred before another.

As the choice of a particular set of constraints is of greater importance in the present application of the general linear model than is usually the case, a large part of the theoretical section in chapter 5 is devoted to the choice of constraints. Some estimation problems are dealt with in the last part of the theoretical section, in particular problems due to insufficient computer capacity. In the second half of chapter 5 some properties of the model are illustrated and evaluated empirically. For this analysis it has been possible to use individual data rather than averages. Different specifications of the general linear model have been applied to several subsets of data. This is done partly in order to economize with computing expenses and partly in order to economize with limited computer capacity. Both restrictions prevent the application of a model with all possible factors and interactions of potential interest and it has been necessary to reduce the size of each problem. This was done both by discarding less important factors and interactions and by applying each model reduced in this way to subsets of data. To this end an investigation was necessary as to which factors and interactions could be omitted without great disadvantage and what effects this procedure would have on the estimates of the remaining parameters. To keep the computing costs low the investigation was partly done by a small model on a very limited amount of data. This may be unsatisfactory from the statistical point of view since the inference drawn is based on a non-probability sample and the judgement that similar results hold for the whole data set.

The second part of chapter 5, i.e. section 5.2, is now organized as follows. The first part explains which models and data sets are used, then follows an empirical illustration of which estimates are obtained with different sets of constraints. In additive models, i.e. models without interactions, comparisons of effects which belong to the same factor are invariant to the three specific sets of constraints used, while this is not in general the case for models with interactions. However, at least for the particular model investigated, the results of such comparisons do not depend critically on the choice between the three sets of constraints.

The investigation of main effects which follows in section 5.2.3 starts with a general survey of the salary structure by an additive model which covers all men with the specified educational qualifications. The effects estimated should properly be interpreted as averages. The main results are that job level is the factor which produces the most differentiation. The average salary at level 2 (the highest level) was in 1968 approximately 2.5 times the average of level 8 (the lowest level) after standardization for differences in age, job family, education, industry and cost of living area. The second most important factor is age. The relative difference between the age classes with the highest and lowest average salaries (factor range) was 90 %. The factor ranges of education and job family were 30 % and 26 % respectively, while cost of living area and industry were the least differentiatory factors, 9 % and 5 % respectively.

The analysis continues with an investigation of the effects of an omitted factor. As the two factors cost of living area and industry are least important as to salary differentiation it is natural to omit them to reduce the size of the model. At least for the particular models investigated this can be done without any important effect on the estimates of the remaining parameters.

The effect due to education referred to above was obtained after standardization for differences in job and other factors. Education, however, determines to some extent the chances of obtaining a high-level job and if this effect is also taken into account, the total effect due to education is higher. One way of obtaining a rough measure of what education means from the point of view of promotion is to estimate the educational effects before and after standardization for job level and to compare. The following may be mentioned as an example of the results obtained. On average a graduate in business & economics received a salary in 1968 which was 8 % higher than that of an employee with high school training in commerce when differences in job attainment were accounted for both being employed in commercial and accounting jobs. Without standardization for job differences, however, the graduate obtained about 35 % more than the non-graduate. The difference, 27 %, can roughly be interpreted as an effect of better promotion which comes to a university graduate.

Similarly the age effects estimated without standardization for differences in job show a higher differentiation than when standardization for job is carried out. The salary increases due to age estimated in chapter 4 include effects of promotion and the corresponding cross section profiles reveal salary differences due to age without standardization for job differences. It should now be possible to get an idea of how much promotion contributes to salary increases if the age effects obtained before standardization for job level are compared with the effects after standardization. Some rough calculations show that measured in this way, promotion contributes almost 2 % to the annual increase of approximately 10 % for employees below 40 years of age and almost nothing above.

If there are interactions among the factors used to explain the salary structure, the additive models only give an »average view» of the salary structure and predictions obtained from this kind of model may contain a considerable error for every subgroup of employees. It is therefore important to investigate the magnitude of possible interactions. This is done in section 5.2.4 with a small model on a limited set of data. The main results are that interactions contribute to salary differentiation by about the same magnitude as the main effects industry and cost of living area. The estimates are, however, rather uncertain and the subset of data analysed small, and this is why the investigation of interactions is pursued in section 5.2.5 as well.

The first sections of the empirical part in chapter 5 which have now been summarized serve the purpose of giving a general survey of the salary structure and a guide as to the possibilities of reducing the size of the models. This information is utilized in section 5.2.5 in which the technical, economic and administrative job families are analysed in some detail. These families cover all major jobs. The technical job families are analysed separately from the non-technical families first by additive models and then also by models with interactions. A comparison between the estimates of the main effects obtained from additive models, applied to technical and non-technical job families, reveals no great differences between these two main groups of families. The results are similar to those obtained before. Omission of the factors industry and cost of living area does not change the estimates of the main effects due to age, education and job very much, nor does the introduction of interactions between these factors. To make the models with interactions manageable the technical job families are divided into two halves and so are the non-technical families. The findings show, among other things, that employees engaged directly in production are better paid than those who work for instance in research and development and, similarly, sales work is better paid than accounting and budget work. The analysis of interactions shows for instance that relative salary differences between graduates and non-graduates decrease by age. This result is obtained after standardization for differences in job attainment and closely agrees with the idea that formal education becomes less important as an employee grows older. A similar result, although uncertain, is obtained for the interaction between education and job level. The relative salary differences between educational groups are smaller at high job levels than at low ones. The most important interaction (by statistical standards) is found between age and job level. The salary difference between high and low job levels increases by age or, to express it in another way, the salary differences between young and old employees are wider at high job levels than at low ones. One possible explanation for this result is that there is a wider scope for experience and skill at high job levels than at low levels.

The data are divided into three non-exclusive groups called »Leavers», »Pairs» and »Beginners», which are supposed to correspond approximately to employees who leave an employer for one reason or another, those who stay with the same employer for at least two consecutive years and those who start on a new job, respectively, The salary structures of these three groups are compared in the last-but-one section of chapter 5. No great differences are observed, but this result is somewhat difficult to interpret as the groups Leavers and Beginners contain a mixture of employees who leave and begin for a number of different reasons and who may also have different salary structures. One conclusion of practical importance is, however, drawn, namely that results obtained from the analysis of Pairs only can be generalized to Leavers and Beginners without great loss of accuracy. Pairs are frequently used both to obtain a more homogeneous set of data and to reduce the number of observations.

The last part of chapter 5 presents the results of an investigation of changes in the salary structure between 1957 and 1968. A general result is that the changes have been rather small. Between 1957 and 1964 relative salary differences due to differences in age decreased somewhat but in 1968 they were again of the same magnitude as in 1957. The differences due to education have hardly changed at all, nor are there any strong indications of changes in differences due to job. A minor decrease in salary differences between industries is, however, observable.

1.3 PRESENT SHORTCOMINGS AND FUTURE RESEARCH

Some of the shortcomings and problems of the methods suggested have already been mentioned and there will be more opportunities in the following chapters to return to this topic. There are only a few concluding remarks.

The profile model in chapter 4 involves an approximation of a continuous profile by a polygon. The purist may find this unsatisfactory, but if the data permit, a closer approximation is easily obtained if the number of segments in the polygon is increased. This approach more readily permits a good fit to various data than enforcement of a particular functional form and the parameters can be conveniently interpreted in a direct way. Some of the assumptions underlying the present model are admittedly rather restrictive. For instance, the initial salary is assumed to follow an exponential function, except for stochastic deviations with zero expectation. As has already been mentioned, a natural development would be to couple the present model with another model which explains the salary increases in economic and/or institutional terms. This is true not only for increases in initial salaries but for all salary increases. One problem in this development is the rather long time series required, as the present salary of one individual depends on all previous increases and decreases obtained. The coupled model would preferably relate salary increases to labour mobility and when this is achieved our understanding of the labour market may improve considerably.

In the present approach with the general linear model it is at least to some extent necessary to specify in advance which factors, categories and combinations of categories are relevant for an explanation of earnings differences. For practical reasons it is impossible to try all possible combinations. For application in situations when the analyser does not know in advance which combinations of factors and categories are most important, Morgan and Sonquist have suggested a method (Morgan and Sonquist [1963]) by which the total data set is sequentially divided into subgroups so that a residual variance is minimized. The advantage of this process is that it is not necessary in advance to fix the analysis of certain combinations of variables, but it is the search procedure which will determine the grouping. An obvious disadvantage is that one may easily obtain a grouping whose interpretation from the point of view of the particular application will be difficult. Another disadvantage is that the sequential procedure does not necessarily give the grouping which corresponds to an unconditional minimum residual variance. A classification analysis according to the suggestion by Morgan and Sonquist could possibly be used as a complement to the kind of analysis made in chapter 5 of this study. It would then be primarily a means of identifying interactions prior to the analysis based on the general linear model. This approach is not used in the present study, however. Instead the choice of interactions for investigation is made on the basis of results from other studies and on an analysis of »small» models applied to subsets of data.

There are also other features of the application of the general linear model which some users may see as disadvantages. It leaves some arbitrariness in the choice of constraints. However, the empirical results show no important sensitivity to the choice between the three sets used in this study. Another problem with this kind of analysis is the relatively high computing expense, but with the present rapid development of computer techniques its share of the total costs is reduced.

Although the approaches suggested in this study for the analysis of earnings data can certainly be improved in various ways, it is the opinion of the author that the empirical results show that they are very promising. A detailed evaluation of possibilities inherent in the methods will of course have to be done for each application separately. This is left for the reader in the following chapters.

CHAPTER 2

INSTITUTIONAL BACKGROUND AND STATISTICAL DATA

2.1 DEFINITION OF THE POPULATION

The statistical data used in this study has been mainly obtained from the salary statistics of the Swedish Employers' Confederation (SAF). Salary statistics from the Swedish Association of Graduate Engineers (CF) are also used to estimate age-earnings profiles. There are no salary statistics of the same quality as those of SAF available in Sweden for a period of consecutive years of the same length. In particular the job classification system used by SAF is unique. Use of this data source limits the population studied to salary earners employed by members of SAF. An additional limitation which is not imposed by the data is that only employees who meet certain educational requirements (see below) will be studied. Defined in this way, what proportion of the total Swedish labour market does the data analysed constitute? According to the population survey in 1960 there were close to 2.3 million men and 1 million women in the labour force.* Of these approximately 1 million men and 250 000 women were employed in mining, manufacturing industry and building and construction. About 20 % of the male (228 077) and 36 % of the female employees (93 338) in these sectors were salary earners while the rest were wage earners.** The educational qualifications at the academic and high school levels most common in Swedish industry are those in engineering, business and commerce.*** The population survey gives the totals for employees with these qualifications in 1960 as 56 474 men and 2684 women. With very few exceptions they are all salary earners. We thus find that they make up 25 % of all male and 3 % of all female salary earners.

Only about 65 % of the male and 50 % of the female salary earners were employed by members of SAF. If we limit our interest to employees with an education in engineering or business and commerce the figures are 61 % and 48 % respectively for 1960. Only about 3 % of the total number of salary earners with these qualifications recorded in the SAF statistics are women,

---

* Employed or temporarily unemployed in at least half time work during the week October 1–8, 1960.

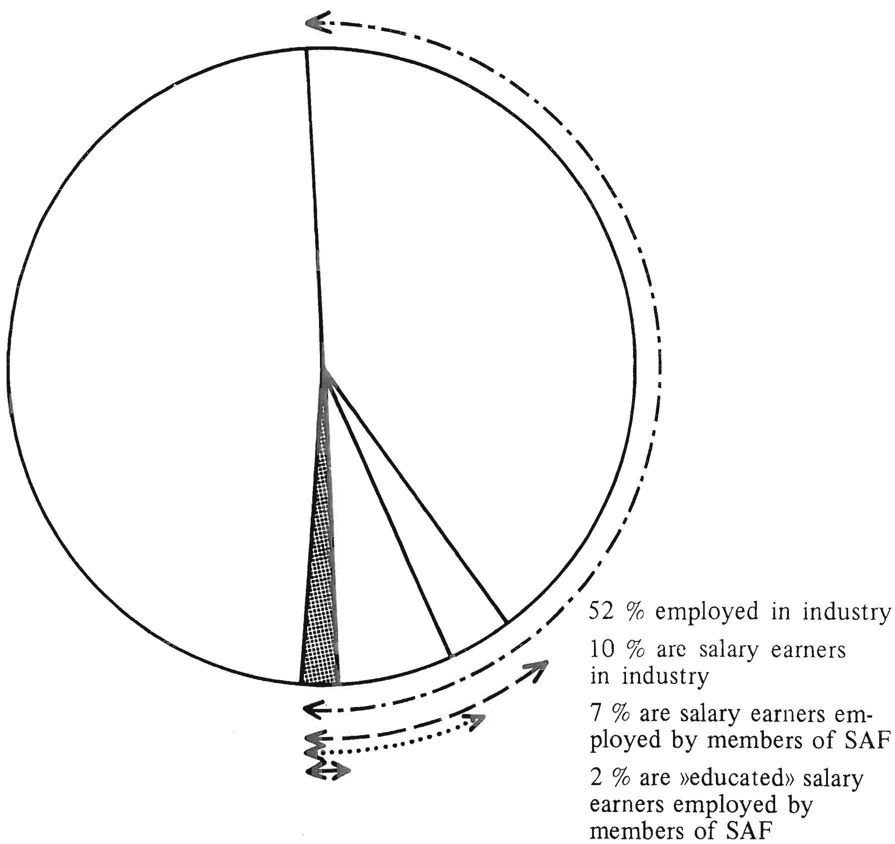
** Salary earners receive monthly salaries while wage earners are paid per hour or per week.

*** Civil-, läroverks- and institutsingenjörer and civil- and gymnasieekonomer.

which is why the study will be limited to men. Figure 2:1 summarizes some of the characteristics mentioned.

In the SAF statistics only academic qualifications and a few vocational high school qualifications are coded while a general high school education is not distinguished from other less advanced educational levels. Of all such salary earners in the SAF sector there are only 4 % with an academic degree. This proportion has stayed approximately constant during the whole period of investigation. The proportion with a (vocational) high school education was 20 % in the mid-fifties and has since then increased a little. In 1967 the proportion was 25 %. Compared with the total number of salary earners in SAF or the labour force as a whole, our population is thus rather small, covering altogether some 40 000 to 50 000 employees.

Figure 2:1. *2.3 million men in the labour force in 1960*



Source: Census of the Population in 1960 and SAF salary statistics 1960.

2.2 INDIVIDUAL SALARY SETTING; CENTRAL AND LOCAL NEGOTIATIONS

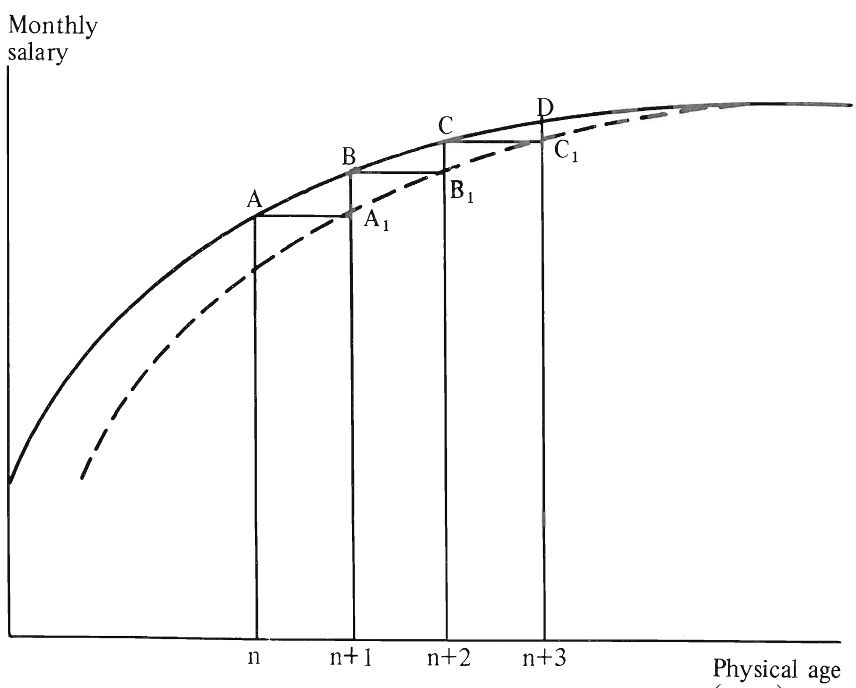
The following is a very brief survey of the salary setting process in the SAF sector of Swedish industry. It is based on Lind [1963], TCO [1966], Eriksson [1968], SAF–SIF [1968], Holmberg [1969], SIF–SALF–CF [1970], and SAF [yearly]. No attempt is made to give a complete picture of the principles underlying the negotiations or of the institutional setting. The interested reader is referred to the references given and to the literature about job and merit evaluation. The purpose of this section is to give the reader who is not familiar with the negotiations for salaried employees in Swedish industry some institutional background which it is hoped will make it easier to understand the great interest in salary comparisons and the instrumental role of the salary statistics in these comparisons.

In most publications about the salary setting process it is emphasized that salary setting is individual and differentiated. Employers and employees have agreed upon three general principles which are basic to salary differentiation. It should be made according to job content, i.e. the skills, responsibility, effort, etc. required for a job; the individual merits and ability to fulfil these requirements and to supply and demand conditions on the market. Job content is defined by a job classification system which is applied both by employers and by employee organizations. The nature of this system is described in the following section. Some attempts have been made to apply formal systems of evaluating merits and ability, but there is no commonly accepted method and practices vary from one enterprise to another. It is probably the most common procedure that no formal system is used at all. Each job is thus classified according to the job classification system, but this does not mean that everyone with the same job code obtains the same salary. Individual merits and ability leave a wide range for salary differentiation within the same job. However, as there is no accepted method of measuring merits and ability it is probably difficult in practice to maintain the distinction between the characteristics of a job and the qualities of an employee. (Nor is this distinction conceptually very clear. An example is that education actually helps to define the job. See below!)

Individual salary setting implies that no formal salary steps or schedules of rates are used. Employers and employees try instead to apply the principles outlined above in a system of central and local negotiations. In the central negotiations between SAF on the one hand and the employee organizations on the other a general framework is agreed upon which takes the form of a recommendation to the local members of the main organizations. This usually includes a general percentage increase for every salary earner and sometimes additional general percentage increases for certain groups, for instance women or employees above a certain age. The central agreements may also contain a recommendation for individual increases, i.e. a certain percentage of the salary sum is reserved for individual allocation in local negotiations. The agreements also include a recommendation that normal age and qualification increments should be added to the negotiated general increase. These increments are not meant to increase the general salary level but are compensations for structural changes within the group of salary earners. It is not possible to give an adequate description of these increments in this short survey, but an attempt will be made to describe how the so called statistical increments due to age are calculated and used.

The solid curve in figure 2:2 represents a cross section profile of average salaries for a particular job or group of jobs. The average salary of employees n years old is A, the average of those who are n+1 years is B, etc. If no increases are given, the same salaries will prevail the next year, while the employees will become one year older. The new cross section profile would then be the broken

Figure 2:2. *The principle of statistical increments due to age*



Source: SAF [yearly].

curve in figure 2:2 and the average salary would decrease. To compensate for this a statistical increment is calculated for each employee, equal to the distance BA_1 for those who are n years old, CB_1 for those who are n+1 years old, etc. The curves and therefore also the increments are calculated from median salaries obtained from the salary statistics each year. For this reason the increments may assume new values each year as the curves change in shape. Each employee does not necessarily obtain »his» increment due to age. Each employer adds up the increments calculated for his employees and the sum is (in principle) redistributed in the local negotiations between the employer and the local unions for salaried employees. The salary increments due to age thus belong to the category of individual increases.

In the local negotiations the central recommendation is implemented and adapted to particular local conditions. Employer and employee representatives give proposals for the allocation of sums reserved for individual allocation. In principle the job and the performance of each individual salary earner are reviewed and his relative salary position evaluated. Any disagreements which may arise are negotiated.

To be more specific we will consider the agreement for 1970. The central recommendation contained the following main points:

a. Each salary is increased by 3 %.
b. 3.3 % of the sum of all salaries (men and women) in each company is reserved for individual distribution.
c. 3 % of the salary sum of all female salary earners in each company is reserved for individual distribution among female salary earners.
d. 2 % of the salary sum for male and female salary earners in each company who were born in 1946 or earlier and have a salary less than Sw.kr. 2000 per month is reserved for individual distribution among these salary earners.
e. 1 % of the salary sum for supervisors at job levels 6 and 7 in each company is reserved for distribution among supervisors who have a low salary compared with the employees they supervise.

According to calculations made in the course of negotiations, these five components will increase total labour costs in the SAF sector by 7 % excluding social costs like pensions and labour taxes. As has been mentioned already the central recommendation also includes stipulations about salary changes which are neutral with respect to the general salary level. The most important are the increases due to age and qualification. From this stipulation and from b – e it follows that there are five »pots» to be distributed among individual salary earners in this agreement. In the local negotiations these pots are calculated for each company and distributed to individuals. This means that everyone obtains at least a 3 %

increase but some obtain much more. For an individual company the increase in labour costs due to this agreement may deviate from 7 % to a greater or lesser extent, but for SAF employers as a whole it is 7 %.

It is easily seen that salary statistics are very important in this whole process. In the central negotiations statistics are used for comparisons mainly between broad groups such as men and women, old and young employees, wage earners and salary earners and between members of different organizations. They are also used to calculate the increased labour cost resulting from alternative constructions of an agreement.

The details in the statistics are even more important at the local level since comparisons are made between individual employees. It is obvious that standardization with respect to differences in age, job and other factors is of great importance. It may be mentioned as an additional example of how the statistics are utilized that each member of SAF receives each year a table from the SAF statistics division which lists the salary records for all his employees and in addition contains with respect to each employee a comparison with the median salary and the first and third quartiles applicable to all employees of the same age, job and cost of living area. On the basis of this table every employer can work out what were the actual results of the previous year's negotiations in his enterprise relative to others and also to what extent he had succeeded in his particular salary policy.

Good statistics are thus very essential in the negotiations and this is also true for the methods used in analysing them.

2.3 DATA COLLECTION AND VARIABLE DEFINITIONS

2.3.1 *The SAF salary statistics*

Once a year, usually in August★ when salary negotiations have been completed and their results implemented locally, salary statistics are collected from each member of SAF in the form of information relating to each full-time salary earner employed. A record is built up for each employee, containing information both about the employer and the employee. The employer is identified by his membership number, which is allocated roughly on the basis of one per enterprise. There are exceptions, however, inasmuch as large concerns may have more than one membership number. If one enterprise has more than one plant, each plant is identified by a plant number. Additional information about the plant (employer) includes employer organization, branch of industry and cost of living area. For each employee there is a job identification, a code for education and information about age and remuneration.

★ Exceptions are May 1961 and September 1966.

The industry and branch of industry nomenclature is the same as that employed by the Central Bureau of Statistics (see below). Before 1962 there were four cost of living areas in Sweden, 2–5. No. 5 is the most expensive area and includes Stockholm and the northern part of Sweden, while No. 2 is the least expensive area. In 1962 areas 2 and 3 were combined into one area called No. 3. To facilitate comparisons the two areas have been analysed together even when a separation could have been made.

A job classification system jointly agreed to by the Swedish Employers' Confederation, the Swedish Union of Clerical and Technical Employees in Industry (SIF) and the Swedish Foremen's and Supervisors' Union (SALF) was first published in 1955 and revised in 1956. There was one major revision in 1965/66 when the Swedish Union of Commercial Employees joined in the collaboration. The following summary of its most important features is based on Lind [1963] and the third edition of the classification system (Position Classification System. Salaried Employees [1968]). There are two dimensions in the system. One dimension is a classification according to job content and the other according to the degree of difficulty or responsibility of the job. Job content is defined by a number of *job families* which are grouped together into ten *occupational fields.* (Before the last revision there were only nine fields, the job families in field 0 being distributed among 1–9.) The ten fields are

0. Administrative work
1. Management and supervision of production
2. Research, experimental and development work
3. Design engineering, industrial and fashion design
4. Other technical work
5. Humanistic and artistic work
6. Teaching
7. General service and health care
8. Commercial work
9. Financial and office work.

An example of a job family in field 1 is Supervision of production, Maintenance, Transport and repair work (110), and another example from field 8 is Marketing (810).

The second dimension of the classification is called *job level.* Each family is divided into a maximum of seven job levels, 2–8. Job level is determined by the responsibility and difficulty of the job. Level 2 indicates the most responsible and difficult job and level 8 the least responsible and difficult one. As an example we may consider the job family Supervision of production, Maintenance, Transport and repair works which contains five levels. They are

2. Chief production manager
3. Production manager
4. First production engineer
5. Production engineer
6. First production technician.

A job level is usually determined by the number and position of employees supervised. This rule is, however, not implemented in the same way in all job families. For instance, a Chief production manager whose job is classified at level 2 has the responsibility for 250–1250 workers while a Chief calculator whose job is at the same level is assisted by two to five subordinates at level 3. Job levels are thus not exactly comparable between families. Another important aspect of job level classification is the education normally required. To quote the nomenclature (pp. 9–10):

»The definitions do not normally state what qualifications and experience are consistent with the duties described. These requirements, moreover, vary between the job families. But for jobs which do not entail administrative and/or supervisory responsibility, the following table is broadly applicable:

Job category with level figure:	Duties normally entail *either* theoretical knowledge corresponding to education as mentioned below, *or* comparable practical training:
2–4	University education
4–6	Gymnasium or technical high school education
6–8	Comprehensive school education (9 years)

The table must, however, be applied with discretion, bearing in mind that the classification of a job is always to be determined on the basis of the definition for the respective job category in the system. Hence, it is apparent that a job can be placed in level 4, for example, even if the person who holds the position has not had a university, gymnasium or technical high school education.»

When in chapter 5 we consider salary differentiation due to education and job level this association between the two variables will be found very important for the results.

The codes in the system are four-digit numbers. The three first digits identify the job family; the first of these digits identifies the occupational field. When a three-digit job family is combined with a job level digit a *job type* or a *job category* or briefly a *job* is identified by the four digits. There are approximately 250 jobs classified in the system each being separately defined by a description in the nomenclature. It is recommended that application of the system to a particular job should be done in cooperation between employer and the local SIF or SALF union, but it is not one of the matters which are negotiable. If a local agreement cannot be reached the matter is submitted to the central organizations. An agreement is reached in most cases but if this is not the case, each party classifies according to his interpretation of the nomenclature. The classification for the SAF statistics is thus done by the employers.

Owing to the revision of the classification system in 1965/66, it is difficult to compare the job structure and the salary structure before and after the revision. An investigation of so called identical individuals made by SAF but not published gives some indication of the results of the revision. An individual is classified as identical if he is recorded during two consecutive years, for instance 1965 and 1966 irrespective of whether or not he moved between employers or organizations. Table 2:1 shows the change in classification and some aspects of the observed differences between 1965 and 1966. The total mobility between job families and job levels is, however, not only a result of the revision, since there is always a more or less »normal» mobility even without a revision. A similar comparison between 1958 and 1959, when the same classification was used in both years, shows that 87 % of all identical salary earners kept the same job code and that 7 % experienced a promotion to a higher level. The corresponding figures for 1965/66 are 65 % and about 15 %. The proportion of employees who were moved to a lower level was about 3 % on both occasions. Samples from 1964/65, when there was no revision, give approximately the same result as those for 1958/59. Some job families were not formally affected by the revision, for instance Mathematical work (200) and Laboratory work (210), but some results from the investigation indicate that these families were also checked and the classification of jobs revised. The result was an unusually large upgrading. The general effect of the revision was therefore not only the introduction of new job families belonging to the field Administrative work (0) and a new distribution among families but also an upgrading of job levels.

The investigation referred to covers all salary earners (table 2:1 only gives statistics for men) while this study is limited to salary earners with certain educational qualifications. As there is no particular investigation of the classification of this group no immediate inference can be drawn, but on the other hand no reasons have been found why the general tendency should not hold.

There are nine educational qualifications coded in the SAF statistics.★ They are

1. Degree in engineering from a university (Civ.ing., Tekn.lic., Tekn.dr.). This category is usually denoted *graduate engineer* in the following.
2. Certificate from a technical high school (Läroverksingenjör). This education usually takes 3 years after 9 years of comprehensive school, but there are also other combinations. Notation: *certificate in engineering I.*

★ In 1969 a more detailed classification was introduced.

Table 2:1. *Changes in the classification of identical male salary earners at the introduction of the third edition of the job classification system in 1965/66*

Family 1966	Number of identical salary earners	Family 1965	Percentage distribution 1965: Higher level 1965 than 1966	Same level 1965 as 1966	Lower level 1965 than 1966	Other families 1965
110	8 020	110+490	6	56	12	26
120	29 386	120+130	10	64	21	5
140	6 911	120+130	5	59	31	5
200	1 330	200	2	66	11	21
210	8 184	210	3	75	15	7
310	13 483	310	3	67	17	13
320	1 463	310+330	3	49	21	27
400	6 594	400+405	5	66	14	15
410	4 139	410	6	62	6	26
440	2 344	440	6	65	6	23
470	3 320	470	6	63	7	24
Totals	85 174		6	65	17	12
810	9 496	810+9006	9	60	6	25
820	11 139	820+9006	8	71	8	13
840	1 861	840+9006	7	49	4	40
850	1 794	850+9006	5	54	1	40
860	661	860	9	66	5	20
870	3 222	870+9006	10	63	6	21
Totals	28 173		8	64	6	22
900	3 197	900	5	71	17	7
910	2 857	910+930	3	58	21	18
962	139	960	1	28	53	18
964	830	960	1	42	29	28
966	885	960	6	59	24	11
990	20 861	900	8	70	11	11
Totals	28 769		7	67	14	12
			All levels			
010	479	110+810+820		63		37
020	510	fields 2–4 and 810+900+910		88		12
040	899	780+790		71		29
060	76	7103		94		6
080	104	560+860		74		26
Total	2 068					

Unpublished source: U. Andrén: Förändringar av de identiska tjänstemännens befattningsklassificering vid införandet av den tredje upplagan av befattningsnomenklaturen. SAF 8.8.67.

3. Certificate from a technical institute (Institutsingenjör). This education does not give the same broad technical education as 2 above. It usually attracts students with a more practical bent and normally extends for 1–2 years. Notation: *certificate in engineering II.*
4. Degree from a business school (Civ.ekon., Ekon.lic., Ekon.dr.). Notation: *degree in business & economics.*
5. High school certificate in commerce (Gymnasieekonom). To obtain this certificate 3 years of study are normally required after 9 years of comprehensive school. Notation: *certificate in commerce.*
6. Degree in social work (Socionom). 3–4 years in college of social work which also includes some practice in the field. Notation: *degree in social work.*
7. *Degree in science* (Fil.kand., Fil.mag., Fil.lic., Fil.dr.). Same notation is used.
8. Degree in law or the social sciences (Jur.kand., Jur.lic., Jur.dr., Pol.mag., and Fil.kand., Fil.mag., Fil.lic., Fil.dr. in the social sciences). The social sciences are for instance economics, business economics (marketing, accounting, management), statistics (but not mathematical statistics), computer sciences and information processing, political science, sociology, psychology and pedagogics. As some of these subjects also form part of other educational qualifications there is some possibility of substitution, in particular between 4, 7 and 8. Notation: *degree in law or social sciences.*
9. Other university degrees. Degrees in medicine, humanities, theology, farmacy, forestry or agriculture. Notation: *other degrees.*

No changes of any importance as regards these nine educational qualifications have taken place during the sample period. The educational variable thus provides a consistent grouping.

The salary concept used is what may be called a gross monthly salary. It includes the stipulated basic rate during the month of investigation (usually August). No deductions have been made for absence from work or individual payments to pension funds or taxes, nor has overtime pay or the employers' part of payments to pension funds been added or any adjustments made in respect of differences in hours of work. The total salary also includes an estimated average payment in kind, performance bonus and shift differential per month during the year. These additions are usually a very small proportion of gross salaries.

The statistical data are stored on tape each year in such a way that it is very expensive to follow identical individuals through time. The combined time series and cross section analyses in chapter 4 have therefore been based on the median salaries published (SAF [yearly]). However, for some applications the data sets for two consecutive years have been matched to form three groups of salary earners. The *first* group includes employees who are recorded with the same employer (membership number) for both years. They are called »Pairs». The *second* group includes those who are recorded with one employer the first year but not with the same employer the second year. This group is called »Leavers». To this group belong employees who leave one SAF member and move to another as well as those who move to public service or other non-SAF members or leave the labour market. Because the SAF statistics do not cover employees at management level (see below), this group also includes all those who are promoted to a management position. The *third* group is called »Beginners» and includes those who were recorded with a SAF member for the second year but not with the same member for the first year. This group covers salary earners who move from one SAF member to another (a subset which intersects Leavers), but it also includes those who move from non-SAF members and those who come directly from schools and universities. Some employees who may have worked for the same employer both years, namely wage earners who obtain promotion to the job of work supervisor are also included among the Beginners. As wage earners are not covered by the salary statistics they are registered as Beginners. Although there may be reasons other than those mentioned why an employee is not recorded. Leavers approximately correspond to what is normally denoted quitters and Beginners approximately correspond to those who take a new job.

The SAF salary statistics are collected from all members of SAF and are not obtained by a sample survey. The response as given in table 2:2 is measured relative to the number of salaried employees in the SAF member record less employees at top management levels who are not included in the survey. The figures in table 2:2 do not directly refer to non-response by education code, but they do indicate that non-response, as is now defined, is no serious problem.

Table 2:2 also shows the increased coverage of the SAF statistics. The number of employees almost doubled between 1956 and 1968. To some extent it pictures a real increase in the number of salary earners in Swedish Industry as a whole, but there is also an effect of SAF's efforts to secure new members. It was too expensive to investigate all new members and all members who leave SAF each year, but two years 1958 and 1964, were selected for investigation. For both years almost all companies applying for membership were small. Almost 50 % had no salary earners at all and of the remaining 50 % hardly any had more than ten. Although this is exceptional it may come about that a big corporation joins SAF. Membership applications as well as withdrawals are most frequent from bakeries and from common carriers. These small enterprises do not usually employ salary earners with an advanced education, which indicates that this study of educated salary earners is less disturbed by SAF's attempts to secure new members than a study of all salary

Table 2:2. *Response to the SAF salary statistics in 1956–1968*

Month	Year	Responding salary earners		Response rate (%)
		Men	Women	Men and women
August	1956	113 638	37 304	89
August	1957	117 641	38 389	90
August	1958	124 242	40 019	90
August	1959	128 401	41 038	91
August	1960	136 428	44 521	93
May	1961	143 822	46 503	92
August	1962	151 176	48 602	90
August	1963	163 533	51 552	92
August	1964	170 095	53 382	93
August	1965	179 618	55 575	94
September	1966	196 335	64 464	90
August	1967	205 097	64 887	92
August	1968	205 452	63 360	94

Source: SAF [yearly], August 1968.

Note: The response rates are calculated as the total number of salary earners reported to the salary statistics relative to the total number of salary earners registered in the SAF membership record.

earners would be. It is not possible however to disregard completely the effect of increasing SAF membership, in particular when the labour composition at the beginning of the sample period is compared with that at the end.

A copy of SAF's salary statistics is delivered each year to the Central Bureau of Statistics which collects salary statistics from non-SAF members. The combined SAF-SCB statistics cover a few additional percentages of educated salary earners, but as long as the set of data studied is defined by the job code there is no difference in population, since the SCB addition does not have this code. The combined statistics have been used in some analyses for practical reasons.

It has already been mentioned that employees at top management level are not covered by the SAF salary statistics. This is a considerable disadvantage as regards the study of age-earnings profiles because it results in an underestimation of the salary increases of middle aged and old employees. Those who are to be promoted to the management level contribute to the average salary increase for young employees as long as they remain in the statistics, but the observed increases of middle aged and older employees are only based on those who were not qualified for further promotion to management positions. There are no adequate salary statistics which cover the top management level and it is thus impossible to estimate the size of this selection bias, but a comparison has been made with statistics collected by the Swedish Association of Graduate Engineers (CF) to obtain at least some information about it. Before the results from this investigation are summarized it is necessary to explain some characteristic features of the salary statistics compiled by CF.

2.3.2 *The salary statistics of the Swedish Association of Graduate Engineers as compared with the SAF statistics*

The Swedish Association of Graduate Engineers (CF) is organized in sections defined by educational specialization. Each section produces salary statistics covering not only members of the association but also non-members with the same specialization. In this study statistics from all sections will be used to evaluate the selection bias in the SAF statistics; in chapter 4 statistics of engineers specializing in electrical engineering (SER) and mechanical engineering (SMR) will be used to estimate age-earnings profiles.★

Once a year (usually in August) CF mails its questionnaire to members and non-members. Free enterprisers and engineers abroad are, however, excluded. A register of non-members is devised from the graduation lists obtained from the colleges of technology. It is, of course, difficult to keep this register up to date and very likely many middle aged and older engineers are missing. As the response rate is very much lower among old than among young engineers, some sections do not collect statistics or at least do not present the results for engineers older than approximately 50 years. Table 2:3 gives the overall response rate for SER and SMR.

Table 2:3. *Response rate to the SER and SMR salary statistics in 1961–1970*

Year	Response rate (%)	
	SER	SMR
1961	90	80
1962	90	81
1963	90	84
1964	91	86
1965	87	76
1966	87	78
1967	89	76
1968	84	81
1969	84	91
1970	84	94

Source: Salary statistics from SER and SMR.

Note: Response rate is defined as number of questionnaires returned to the total number sent out.

★ SER = Svenska Elektroingenjörers Riksförening
SMR = Svenska Mekanisters Riksförening.

Some of the variations may depend on how well the register of non-members is kept up. The response rate is generally higher for members than for non-members. For each engineer information is collected about his monthly salary, his age and graduation year. The same definitions are applied as in the SAF statistics. It is also noted whether the engineer is employed in public service or by a private employer. The CF statistics do not therefore cover only SAF members.

Although there is a high non-response among middle aged and older employees, there is no reason to believe that all engineers at the top management level belong to non-respondents. We do not know with what frequency this group answer the CF questionnaire, nor is it possible to find from the records since there is no job code, but by a comparison between the CF and SAF statistics for graduate engineers it may be possible to extract some valuable information. To this end a match of individual employees between the two data sources was attempted for 1970.★ The comparison was limited to enterprises which could be identified in both sources. CF does not give an enterprise or company code to small enterprises and companies (size is measured by the total number of employees, no rigid rule being followed) which means that as many as 1738 graduate engineers are excluded from the comparison simply because no code is found in the CF data. In addition another group of approximately 800 engineers is excluded because their employers are not members of SAF. Consultants belong to this category. The identification of *individuals* was therefore attempted for only 6657 graduate engineers. Besides company code the identification was made by date of birth and salary. As social security number is not available in the CF record, the identification cannot be made with complete certainty. 2747 engineers were identified in both registers, the $CF \cap SAF$ group; 1237 were only found in the CF register and 2673 only in the SAF register. The unmatched groups are larger than is commonly claimed by CF and SAF, but there are several reasons for the belief that this is due to the insufficient identification.★★

Despite these limitations table 2:4 shows that middle aged and older engineers are more frequent in the CF and SAF groups than in the $CF \cap SAF$ group. As the coverage and response are not high for these age groups in the CF statistics, this

---

★ This investigation was carried out by Fil.kand. Kurt-Allan Wallgren.

★★ The membership numbers and the plant numbers in SAF do not exactly coincide with the company code used by CF. Although attention was paid to this problem, it is possible that for instance a plant belonging to an enterprise with a certain number in SAF may not be found under the code which roughly corresponds to this enterprise in CF, because the plant there was classified differently, or vice versa. Furthermore, the salary figures do not always coincide exactly, for instance because payment in kind has been valued differently. A certain difference was accepted for identification but nevertheless this error possibility very likely gives too few identical engineers. Although less likely, it is possible that errors in the birth date will also work in the same direction.

result may have been predicted for the SAF group, and for the CF group the result is consistent with the hypothesis that the group covers the top management level. Nevertheless 45 % of the group is younger than 35 years.

Table 2:5 and figure 2:3 show that the monthly mean salary of CF∩SAF and SAF is similar across age groups, while the mean salary of the CF group is constantly higher. The difference increases by age until the interval 50-59 years. The increase is highest around 35 years which may indicate that the engineers above 35 years possess some characteristic different from those below 35, for instance employment at top management level. It is difficult to see why the shortages in the identification of companies should work in any particular direction. The lack of identification because of different salaries may be more frequent among middle aged and old engineers since it may be more common among those who have jobs with larger proportions of fringe benefits. It is however hard to believe that this could explain any substantial part of the difference.

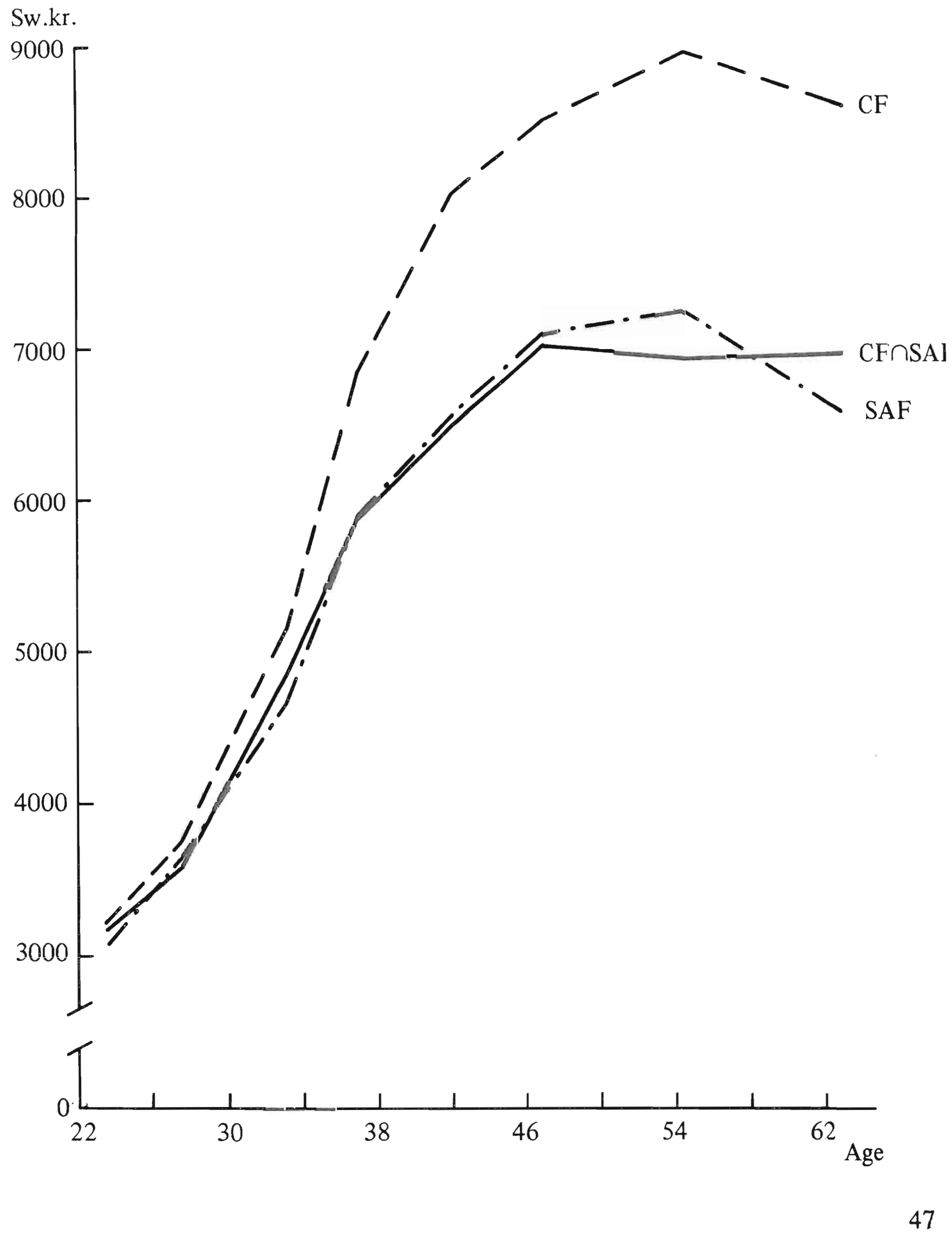
Table 2:4. *Age distribution of the CF∩SAF, CF and SAF groups in 1970 (%)*

Group	Age								All groups	
	22–25	26–29	30–34	35–39	40–44	45–49	50–59	60–	%	Number
CF∩SAF	4.2	28.3	26.2	15.4	11.4	8.9	5.1	0.5	100.0	2747
CF	5.8	18.5	20.6	14.0	15.4	13.5	11.0	1.1	99.9	1237
SAF	2.8	17.7	21.0	14.8	14.6	10.3	14.4	4.4	100.0	2673

Table 2:5. *The CF and SAF mean salary relative to the CF∩SAF mean salary in 1970 (percentage deviations)*

Group	Age							
	22–25	26–29	30–34	35–39	40–44	45–49	50–59	60–
CF	1.0	4.0	6.5	16.4	23.0	21.1	28.9	23.5
SAF	−3.3	0.8	−3.9	−0.3	0.9	1.2	4.7	−5.8

Figure 2:3. *Arithmetic mean salary by age in 1970 (Sw.kr. per month)*



CHAPTER 3

LABOUR COMPOSITION AND MOBILITY

The analysis performed in this chapter of age distribution, educational composition and the employment in jobs and industries will serve as a general background for the analysis of the salary structure in chapters 4 and 5. The interest will be concentrated on the educational composition in different sectors of the labour market such as industries and job families. To describe changes in composition some descriptive statistics will be defined and also applied to the salary statistics from the Swedish Employers' Confederation. The purpose is mainly to describe observable changes rather than to explain them. The properties of the data used have already been analysed in chapter 2. Definitions of variables are also given there.

3.1 DESCRIPTIVE METHODS FOR THE ANALYSIS OF CHANGES IN COMPOSITION

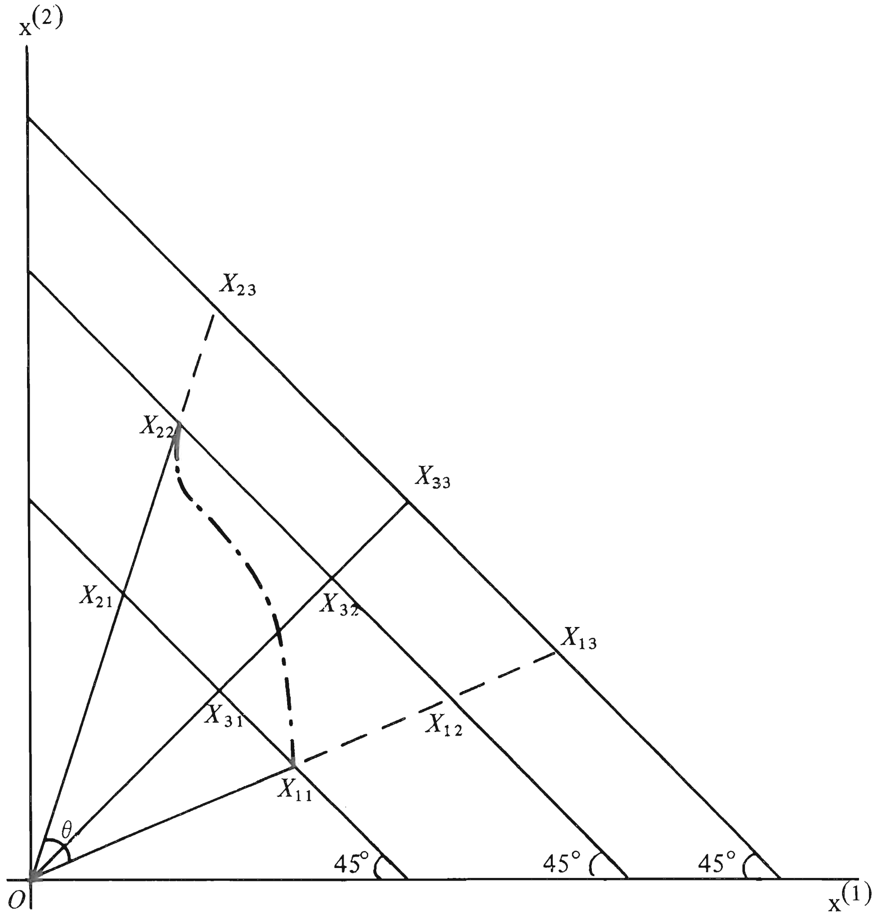
Our methodological problem is to survey the distribution of salaried employees by age, education, job and industry, and to note how this distribution changes from one year to another. Since educational qualifications, jobs and industries are rather numerous we have to look for some relatively simple descriptive statistics which summarize main characteristics, in order to make the survey manageable.

Each sector and point in time is characterized by a certain number of employees distributed in a number of subsectors. In an industry, for instance, there are a number of employees distributed over different educational qualifications and different jobs. A change between two points in time can be described as a change in the total number of employees of the sector and a change in the distribution within subsectors.

Suppose there is a group of employees where only two educational qualifications are registered, (1) and (2). Suppose also that the group composition at time 1 is represented by $X_{11} = (x_{11}^{(1)}, x_{11}^{(2)})$ in figure 3:1 and the composition at time 2 by $X_{22} = (x_{22}^{(1)}, x_{22}^{(2)})$. Between these two points in time the total number of employees as well as the educational mix have changed. The latter implies that X_{22} does not lie on the straight line OX_{11} past X_{11}. These two changes will in the following be called change in level and change in mix.*

* »Level» used in this context should be distinguished from job level.

Figure 3:1. *Geometrical interpretation of measures of changes in level and mix*



It is not known what changes have taken place between the two observed points in time. The movement from the first observed point to the second point may for instance have followed the chain line between the points X_{11} and X_{22} in figure 3:1. Since we do not know the exact path followed, we may attempt to describe the changes by two straight lines. One passes through X_{11} and the other through X_{22} ; both at an angle of 45° with the axes. The first line is the locus of all combinations of employees with the two educational qualifications which total $x_{11}^{(1)}+$ $+ x_{11}^{(2)}$ and the second is the locus of all combinations which total $x_{22}^{(1)} + x_{22}^{(2)}$ employees. X_{22} can now be reached from X_{11} in two different ways, firstly by way of X_{12} and secondly by way of X_{21}. (To understand the subindex notation of X we observe that X_{ij} is a point at level j on the radius from the origin determined by the *observed* point X_{ii}.) In the first case the change in level can be measured by the vector $X_{11}X_{12}$ and the change in mix by $X_{12}X_{22}$ and in the second case the measures are $X_{21}X_{22}$ and $X_{11}X_{21}$ respectively. These vectors will be called level vector and mix vector.

The vector distances may be used as a more compact measure of level changes and mix changes, but as there are two alternative vectors describing each change, there is no unique distance measure, nor is there a natural preference order between them. In general $|X_{11}X_{12}| \neq |X_{21}X_{22}|$ and $|X_{11}X_{21}| \neq |X_{12}X_{22}|$. As uniqueness is a desirable property it is better to use either of the following equally large ratios as a measure of changes in level

$$Q_{12} = \frac{|OX_{12}|}{|OX_{11}|} = \frac{|OX_{22}|}{|OX_{21}|}; \tag{3:1}$$

Q_{12} minus one multiplied by one hundred is nothing but the percentage rate of increase in the total number of employees.

There is also another desirable property which the descriptive statistics should possess. To describe changes which have taken place between more than two points in time it is desirable that the statistics for the subperiods should be consistent with the statistic for the whole period. It should be possible to follow some simple rule and link together the statistics for the subperiods to describe the total change which should then give the same result as when the same statistic is applied to the whole period. For the ratio (3:1) the rule is obviously a simple multiplication.

$$Q_{13} = Q_{12}Q_{23}; \tag{3:2}$$

Index numbers are usually required to satisfy the so called circular criterion, and the same requirement can be imposed on Q. It simply means that the principle used to link the Q's should give the original value of Q, say $Q_{11} = 1$, if the changes in level after a number of steps bring the total number of employees back to the first level, for instance

$$Q_{11} = Q_{12} \cdot Q_{23} \cdot Q_{31}; \tag{3:3}$$

Note that (3:3) does not imply that the first and last points are the same. Although the total number of employees is the same, the mix does not have to be the same. The truth of (3:3) is obvious and no proof is necessary. It is also easily seen that Q satisfies the circular criterion for any number of steps.

To use the distances of the two alternative mix vectors $X_{11}X_{21}$ and $X_{12}X_{22}$ as measures of the change in mix has the disadvantage that there is no criterion of determining the choice between them. There is also the disadvantage that they both depend on the level. It is possible to avoid these two drawbacks if the change in mix is measured by the angle θ between OX_{11} and OX_{22}. θ, however, has other undesirable properties. For a given distance $|X_{12}X_{22}|$ θ will take a smaller value if X_{12} and X_{22} are close to one of the axes than if they are not. Furthermore, if there are only two educational qualifications it is possible to define a unique positive direction of change, for instance a change parallel to the 45° lines towards one of the axes may be defined as the positive direction while a change in the opposite direction towards the other axis is the negative direction. To link successive θ's the principle to follow would be a simple summation. If there are more than two educational qualifications, however, the problem cannot be handled so simply. The same angle can then be obtained from completely different changes in mix and a unique measure requires a statement of more than one direction. The simplicity of the measure is then lost.

Another measure of changes in mix can be obtained if the 45° lines are moved towards the origin to satisfy the points (1, 0) and (0, 1) and if the mix vector is defined along this new 45° line. This is the same thing as to calculate the relative educational composition

$$H_t = \{\eta_t^{(1)}, \eta_t^{(2)}\} = \left\{\frac{x_{tt}^{(1)}}{x_{tt}^{(1)} + x_{tt}^{(2)}}, \frac{x_{tt}^{(2)}}{x_{tt}^{(1)} + x_{tt}^{(2)}}\right\}; \quad t=1,2,\ldots \tag{3:4}$$

and then to apply the definition of a mix vector to (3:4).

$$|H_1H_2| = \sqrt{(\eta_2^{(1)} - \eta_1^{(1)})^2 + (\eta_2^{(2)} - \eta_1^{(2)})^2}; \tag{3:5}$$

(3:5) can be generalized to cases with more than two educational qualifications.

$$|H_tH_{t+k}| = \sqrt{\sum_{i=1}^{n} (\eta_{t+k}^{(i)} - \eta_t^{(i)})^2}; \tag{3:6}$$

For every value of n, this distance attains a maximum of $\sqrt{2}$ and a minimum of 0. If desired, a standardization is thus easily accomplished if the vector distance is divided by $\sqrt{2}$. We may also note that the mean of the differences $(\eta_{t+k}^i - \eta_t^i)$ is zero since the sum of all η^i is always equal to one. If the vector distance (3:6) is divided by $\sqrt{n}$, we obtain a measure which can be interpreted as a standard deviation.

$$S_{t,\ t+k} = \frac{|H_tH_{t+k}|}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n} (\eta_{t+k}^{(i)} - \eta_t^{(i)})^2}{n}}\ ; \qquad (3:7)$$

As the maximum vector distance is invariant of n, max S will decrease as n increases. An S based on n-dimensional data does not have to be smaller than S based on m-dimensional data, m < n, but one intuitively understands that this will usually be the case. The S measure is thus unsuitable for comparisons of two mix changes which take place in spaces of different dimensions. Although it is not clear that a comparison of a mix change in n dimensions with a change in m dimensions is meaningful, it is suggested that $S\sqrt{n}$ or $S\sqrt{\frac{n}{2}}$ is a better measure than S in this case.

Which rule should in this case be followed to link successive S's? Suppose there are observations from three different occasions, 1, 2 and 3. S_{13}^2 can be broken down in the following way.

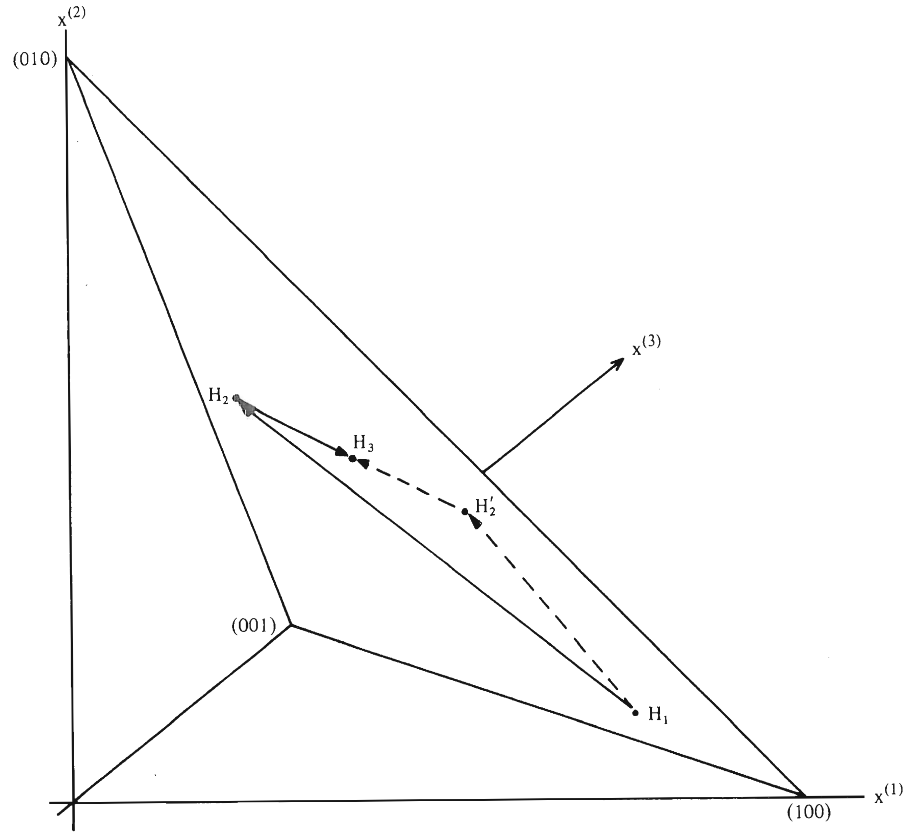
$$S_{13}^2 = \frac{\sum_{i=1}^{n} (\eta_3^{(i)} - \eta_1^{(i)})^2}{n} = \frac{\sum_{i=1}^{n} (\eta_3^{(i)} - \eta_2^{(i)})^2}{n} + \frac{\sum_{i=1}^{n} (\eta_2^{(i)} - \eta_1^{(i)})^2}{n} + \frac{2\sum_{i=1}^{n} (\eta_3^{(i)} - \eta_2^{(i)})(\eta_2^{(i)} - \eta_1^{(i)})}{n} \qquad (3:8)$$

and in a more compact notation

$$S_{13}^2 = S_{23}^2 + S_{12}^2 + 2R_{12,23}S_{23}S_{12}; \qquad (3:9)$$

The rule which prescribes how the measures of mix changes will be linked is thus more complicated than the rule for the level ratio, and yet it allows a very intuitive interpretation. The changes in mix between occasion 1 and occasion 3 equal the sum of the changes between 1 and 2 and between 2 and 3 adjusted for differences in the direction of the changes. $R_{12,23}$ is not only the correlation coefficient between the relative changes in mix, but also the cosine of the angle between the two mix vectors H_1H_2 and H_2H_3. Consider for instance the four points H_1, H_2, H_2' and H_3 in the plane (hyper plane) which satisfy the points (1,0, ...0), (0, 1, 0,...0), ...(0, ..., 0, 1) and thus form 45° angles with the axes, (figure 3:2). If the mix changes from H_1 to H_2 and then to H_3, the second change almost goes in a direction opposite to that of the first change which is to some extent neutralized. The sum of the distances between H_1, H_2 and H_2, H_3 will then exaggerate the total change unless an allowance is made for the reversal of direction. If the changes take place instead via H_2' they work in approximately the same direction and the third term in (3:9) will give a positive contribution to S_{13}^2. To see that this agrees with one's intuitive thinking, suppose $R_{12,23}=1$, i.e. the two changes work in exactly the same direction. (3:9) now reduces to

Figure 3:2. *Geometrical interpretation of changes in mix*



$$S_{13}^2 = S_{23}^2 + S_{12}^2 + 2S_{23}S_{12} = (S_{23} + S_{12})^2; \qquad (3:10)$$

Thus

$$S_{13} = S_{23} + S_{12}; \qquad (3:11)$$

i.e. the total change in mix (distance) equals the sum of the two subchanges (distances).

The rule (3:9) and the special case (3:10) can be generalized to more than three points in time. The expression for the total change will then contain correlation coefficients between all subchanges. For instance, with four points in time we obtain

$$S_{14}^2 = S_{12}^2 + S_{23}^2 + S_{34}^2 + 2R_{12,23}S_{12}S_{23} + 2R_{12,34}S_{12}S_{34} + 2R_{23,34}S_{23}S_{34}; \tag{3:12}$$

The correlation coefficient is of interest not only in connection with the S measure but also in its own right. It will be used not only to compare the directions of mix vectors but also to compare the composition of groups in general.

The S measure is reversible and satisfies the circular criterion. By definition the following relation holds

$$S_{12} = S_{21}; \tag{3:13}$$

From (3:9) and (3:13) the resultant change in mix from a movement from one point to another and back again is

$$S_{11}^2 = S_{12}^2 + S_{21}^2 - 2S_{12}S_{21} = 0; \tag{3:14}$$

If no change at all is defined as $S_{11} = 0$, the S measure obviously satisfies the circular criterion in this simple case. For the case with three steps (time periods) (3:12) gives the expression

$$S_{11}^2 = S_{12}^2 + S_{23}^2 + S_{31}^2 + 2R_{12,23}S_{12}S_{23} + 2R_{12,31}S_{12}S_{31} + \\ + 2R_{23,31}S_{23}S_{31}; \tag{3:15}$$

Adding and subtracting the same terms give

$$S_{11}^2 = S_{12}^2 + S_{23}^2 + 2R_{12,23}S_{12}S_{23} + S_{12}^2 + S_{31}^2 + 2R_{12,31}S_{12}S_{31} + \\ + S_{23}^2 + S_{31}^2 + 2R_{23,31}S_{23}S_{31} - (S_{12}^2 + S_{23}^2 + S_{31}^2); \tag{3:16}$$

After substitution of (3:9) into (3:16) and application of (3:13) we finally obtain

$$S_{11}^2 = S_{13}^2 + S_{32}^2 + S_{21}^2 - (S_{12}^2 + S_{23}^2 + S_{31}^2) = 0; \tag{3:17}$$

The circular criterion is thus satisfied and a generalization to any number of steps is obvious.

In the applications which follow below, a point X is characterized by the sum of the elements of the vector OX, i.e. the total number of employees, and by all elements of the corresponding OH vector, i.e. a relative distribution. A change in composition from one point to another is described by the Q and S measures and sometimes also by the difference between the two corresponding OH vectors (relative distributions).*

3.2 EMPLOYMENT BY SECTORS

In 1957 there were 30 140 salaried employees with a degree or a certificate employed by members of the Swedish Employers' Confederation and in 1968 this sum had increased by 66 % to 50 130. The increase took place mainly during the first part of the period. Between 1957 and 1964 the number of employees increased by 52 % and between 1964 and 1968 by only 10 %. The following sections will show how the employees are distributed by education, industries and jobs, and how the change in their total number is reproduced in these distributions.

It has already been pointed out, but it deserves to be mentioned again, that the changes registered are changes in the number of employees recorded which implies that the effects of new members entering SAF and old members leaving and the effects of new classification principles are buried in the calculated measures. As the purpose is primarily to describe the data used in the analysis of the salary structure and not to explain changes in the labour market this is no serious disadvantage. In the previous chapter an attempt was made to get some information as to the magnitudes of these effects and the tentative conclusion is that the changes observed in the data, at least in broad groups, are not influenced very much by these »artificial» effects.

---

* An alternative approach was also considered. Not only changes in composition but also the composition itself can be characterized by summary measures. One might try to find measures which, when applied to a point X_{11}, could be linked with the Q and S measures to characterize a new point X_{22}. This could for instance be done by one measure of level and one of mix. The total number of employees is a natural measure of level. The mix could be related to some reference point, for instance the point with all elements equal to the reciprocal of the number of dimensions, or to some other point which is convenient in a particular application. The measure might for instance be defined as the vector distance from the reference point to the observed point H. The principle of linking this measure with the S measure is similar to that for S. This whole approach was, however, rejected because it was decided that too much information would be lost in the new descriptive statistics.

3.2.1 *Educational composition by industries*

The increases in employment by education and by industry are given in tables 3:1, 3:2 and 3:3. Note that the data for 1968 are obtained from the statistics published by the Swedish Employers' Confederation, while the figures for 1957 and 1964 are obtained from the salary statistics collected in cooperation between SAF and the Central Bureau of Statistics which cover approximately an additional 2 % of salaried employees with the specified educational qualifications.

The rate of increase of course varies from one industry to another. Iron and metal works, Quarrying and Building and construction doubled their employment, while the Beverage and tobacco industries and the Textile industry decreased theirs by 47 % and 26 % respectively (table 3:1). The observation made for all industries, that most of the increase in employment took place during the first part of the period, also holds for each industry with a few exceptions, for instance Mining, Iron and steel works, Other metal industry and Food industry.

The S measure in table 3:1 shows that the change in educational mix was no greater than about 1 %, 0.6 % for each subperiod. Compared to the 62 % increase in the total number of employees★ the mix may be characterized as stable. The S measures calculated for each industry reveal more variability in some industries, in particular in the Beverage and tobacco industries, in Building and construction and in Wood industry. The correlation coefficient $R_{57/64,\,64/68}$ shows that the changes between 1964 and 1968 in many industries went in almost the opposite direction to the change between 1957 and 1964. Examples are Mining, Shipyards, and Other metal industry. As the changes are generally rather small this may be seen as a result of a stochastic variability rather than a systematic and significant turn.

Table 3:2 exhibits the educational composition in all industries taken together. High school certificate in engineering (I and II) is no doubt the dominating education. The two categories made up 77 % of all educated employees in 1968. As graduate engineers made up 11 %, there was (and still is) an overwhelming dominance of engineers among employees with the educational qualifications considered in Swedish industry. The corresponding distributions for each industry in table 3:3 reveal some rather natural differences in educational mix due to differences in production methods and educational requirements. The proportion of graduate engineers is high in Mining, Shipyards, Manufacture of electrical equipment and in the Pulp and paper industry. In some industries where the proportion of graduate engineers is rather low, but technical knowledge is still required, the proportions of high school engineers are high. Examples

★ See note to table 3:2.

are Manufacture of hardware, Repair works, Other metal industry and Building and construction. In the latter industry the proportion of engineers with certificate II was as high as 57 %. The proportion made up by employees with an education in business, economics and commerce is small in Mining, Metal and engineering industry and in Building and construction but they constitute greater proportion in the other industries. This is particularly true for those who have high school training. Educational qualifications other than technical and economic education can be traced to just a few industries, namely Chemical industry, Beverage and tobacco industries, Food manufacturing industries, Printing and allied industries and Mining. The group Other degrees probably contains a large proportion of pharmacists which may explain the relatively high proportion this group represents in the Chemical, Beverage and Food industries.

The mix vectors in table 3:2 show the change in mix. The resultant change from 1957 to 1968 for all industries is mainly a small decrease in the relative number of graduate engineers and high school engineers II and a corresponding increase of high school engineers I. To find out the extent to which the same change has taken place in each individual industry, the correlation coefficients (r) have been calculated between the mix vector for all industries together and the mix vectors for each industry. Each branch of industry in the Manufacturing and engineering industry is compared with the mix vector of the whole industry. The result is shown in the last column of table 3:1. In the Textile industry the changes in mix have gone in a direction almost opposite to the general direction, i.e. the proportion of employees with a certificate in engineering I has decreased and that of graduate engineers has increased. The association is very low in Iron and steel works, metal plants; Food industry and in Leather, furs and rubber industries. In Iron and steel works, metal plants the proportion of graduate engineers and high school engineers I decreased by 4 percentage units and increased by 5 percentage units respectively over the period 1957–1968 which is the same tendency as the general one, but the proportion of high school engineers II also increased by 6 percentage units and that of employees with a certificate in commerce decreased by 3 percentage units which is contrary to the general tendency. The proportion of high school engineers II also increased by 6 percentages in Food manufacturing industries and 5 percentage units in Leather, furs and rubber industries, and in the latter industry the proportion of employees with commercial education also decreased by 6 percentage units.

The Beverage and tobacco industries were previously known as the industry with the greatest change in mix. It is one of the smallest industries with only 209 educated employees in 1957. Between 1957 and 1964 their number increased by 28 %. The change in mix followed very closely the general tendency. The number of graduate engineers decreased from 28 % to 14 % and high school engineers II from 34 % to 26 %, while high school engineers I increased from

Table 3:1. *Measures of changes in employment level and educational mix in 1957–1968*

Industry	$Q_{57,64}$	$Q_{64,68}$	$Q_{57,68}$	$S_{57,64}$	$S_{64,68}$	$S_{57,68}$	$\sqrt{\frac{n}{2}S_{57,68}}$	$R_{57/64,64/68}$	$r_{57,68}$
All industry	1.498	1.084	1.623	0.006	0.006	0.011	0.023	0.586	1.000
Mining	1.195	1.119	1.337	0.033	0.023	0.020	0.042	−0.802	0.637
Metal and engineering industry	1.509	1.069	1.613	0.006	0.001	0.006	0.013	−0.005	0.922
Iron and steel works; metal plants	1.491	1.404	2.093	0.011	0.021	0.027	0.057	0.365	0.086
Manufacture of hardware	1.412	1.185	1.673	0.013	0.005	0.012	0.025	−0.503	0.555
Engineering works	1.450	0.994	1.441	0.007	0.005	0.007	0.015	−0.405	0.637
Repair works	1.562	0.832	1.300	0.017	0.014	0.024	0.050	0.241	0.753
Shipyards	1.256	1.003	1.259	0.021	0.017	0.011	0.023	−0.858	0.934
Manufacture of electrical equipment	1.537	1.098	1.688	0.006	0.011	0.009	0.019	−0.625	0.799
Other metal industry	1.175	1.241	1.458	0.043	0.025	0.015	0.030	−0.962	0.851
Quarrying; stone, clay and glass products	1.566	1.231	1.928	0.021	0.025	0.032	0.068	−0.044	0.950
Wood industry	1.384	1.157	1.600	0.038	0.017	0.037	0.079	−0.157	0.732
Manufacture of pulp, paper and paper products	1.204	0.907	1.092	0.020	0.007	0.025	0.054	0.848	0.572
Printing and allied industries	–	0.441	–	–	0.056	–	–	–	–
Food manufacturing industries	1.198	1.300	1.557	0.014	0.020	0.030	0.064	0.638	0.222
Beverage and tobacco industries	1.278	0.412	0.526	0.075	0.087	0.146	0.292	0.639	0.582
Textile industry	0.896	0.826	0.740	0.009	0.008	0.012	0.025	0.092	−0.566
Leather, furs and rubber industries	1.474	0.953	1.405	0.017	0.021	0.029	0.062	0.241	0.194
Chemical industry	1.539	1.170	1.801	0.016	0.015	0.020	0.042	−0.163	0.761
Building and construction	1.652	1.298	2.143	0.017	0.031	0.045	0.095	0.779	0.840

Note 1: $r_{57,68}$ is the correlation coefficient between the mix sector for the whole industry (metal and engineering industry) and each subindustry (branch in Metal and engineering industry), calculated for the timeperiod 1957–1968.

Note 2: see note to table 3:2.

Table 3:2. *Educational composition in 1957–1968; all industries*

	Educational qualifications									
	Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce	Degree in social work	Degree in science	Degree in law or in the social sciences	Other academic degrees	All educational qualifications
Number of employees 1957	4 053	9 163	14 116	739	2 162	35	163	159	289	30 879★
Number of employees 1964	5 543	14 247	20 824	1 066	3 553	76	360	181	468	46 318★
Number of employees 1968	5 661	16 124	22 532	1 010	3 524	62	423	252	542	50 130
Relative distribution 1957	0.131	0.297	0.457	0.024	0.070	0.001	0.005	0.005	0,010	1.000
Change in mix 1957–1964	−0.011	0.010	−0.008	−0.001	0.007	0.001	0.003	−0.001	0.000	0.000
Relative distribution 1964	0.120	0.307	0.449	0.023	0.077	0.002	0.008	0.004	0.010	1.000
Change in mix 1964–1968	−0.007	0.015	0.001	−0.003	−0.007	−0.001	0.000	0.001	0.001	0.000
Relative distribution 1968	0.113	0.322	0.450	0.020	0.070	0.001	0.008	0.005	0.011	1.000
Change in mix 1957–1968	−0.018	0.025	−0.007	−0.004	0.000	0.000	0.003	0.000	0.001	0.000

★ The figures for 1957 and 1964 are obtained from the statistics collected by SAF and the Central Bureau of Statistics in cooperation. The totals from the SAF statistics only are 30 140 for 1957 and 45 737 for 1964 and the corresponding Q values are $Q_{57,64} = 1.517$, $Q_{64,68} = 1.096$ and $Q_{57,68}=1.663$

Table 3:3. *Percentage distribution of employees by education and industry in 1968*

Industry	Education: Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce	Degree in social work	Degree in science	Degree in law or in the social sciences	Other academic degrees	All educational qualifications %	All educational qualifications number of employees
All industry	11.3	32.2	44.9	2.0	7.0	0.1	0.8	0.5	1.1	99.9	50 130
Mining	20.1	33.6	30.8	2.7	5.0	0.4	3.4	2.5	1.5	100.0	678
Metal and engineering industry	11.9	33.1	46.5	1.7	5.5	0.1	0.6	0.3	0.3	100.0	32 274
Iron and steel works; metal plants	13.9	34.9	39.9	1.5	7.8	0.1	1.1	0.3	0.4	99.9	4 023
Manufacture of hardware	6.4	35.2	43.3	3.3	10.5	0.1	0.3	0.5	0.4	100.0	1 527
Engineering works	9.9	33.1	48.4	1.8	5.6	0.1	0.5	0.3	0.4	100.1	14 939
Repair works	1.3	18.6	46.2	5.8	27.1	0.3	0.0	0.5	0.3	100.1	377
Shipyards	14.9	26.7	52.4	1.2	3.3	0.1	0.7	0.2	0.5	100.0	1 590
Manufacture of electrical equipment	15.8	33.9	45.4	1.3	2.7	0.1	0.5	0.3	0.1	100.1	9 226
Other metal industry	3.9	32.1	50.7	2.2	10.8	0.0	0.0	0.2	0.2	100.1	592
Quarrying; stone, clay and glass products	10.2	35.1	38.5	2.2	11.6	0.3	1.1	0.7	0.3	100.0	1 573
Wood industry	4.9	32.0	36.8	2.7	20.7	0.1	0.4	0.6	1.9	100.1	701
Manufacture of pulp, paper and paper products	15.2	30.8	27.6	3.3	16.2	0.3	0.5	1.3	4.9	100.1	1 986
Printing and allied industries	1.2	18.2	36.4	14.2	20.6	0.4	2.4	4.3	2.4	100.1	253
Food manufacturing industries	7.1	26.8	26.3	6.7	21.7	1.0	3.6	1.8	5.0	100.0	897
Beverage and tobacco industries	8.2	20.9	13.6	3.6	29.1	0.0	0.0	1.8	22.7	99.9	110
Textile industry	7.4	31.4	35.4	6.4	17.6	0.0	0.4	1.1	0.3	100.0	717
Leather, furs and rubber industries	4.2	39.0	33.1	4.6	18.5	0.0	0.0	0.5	0.2	100.1	590
Chemical industry	12.6	27.9	32.6	3.5	9.9	0.3	4.6	1.3	7.2	99.9	3 178
Building and construction	8.7	30.2	57.0	0.5	3.3	0.0	0.0	0.2	0.1	100.0	7 173

Source: SAF salary statistics 1968.

13 % to 21 % and employees with a certificate in commerce from 12 % to 17 %. The group other academic degrees also increased from 0.5 % to 8.6 %. During the second period the total number of employees decreased by 59 % and at the same time the changes in mix deviated from the general tendency. The proportions of all educational groups decreased except certificate in commerce, +12 percentage units, degree in law or in the social sciences, +1 percentage unit, and other academic degrees, +14 percentage units. The proportion of high school engineers II decreased by almost 13 percentage units.

A general and perhaps trivial conclusion from these comparisons of changes in mix is that on an aggregate level the changes are rather small, the mix is stable, but when the data are broken down into smaller groups covering shorter time periods there are greater changes which, however, seem to behave in a rather erratic manner.

3.2.2 *Composition by job and education*

The tool at our disposal for investigation of the job composition is the position classification system used by the Swedish Employers' Confederation (SAF), the Swedish Union of Clerical and Technical Employees in Industry (SIF), the Swedish Union of Supervisors and Foremen (SALF) and the Swedish Union of Commercial Employees (HTF). The results obtained from comparisons of job composition naturally depend to a great extent on the classification system and its application. It was mentioned in the second chapter that the classification system was revised between 1965 and 1966, and this makes a comparison between job compositions before and after the revision very difficult. As the revision was directed more towards the job families than towards the levels, no comparison is made between the job family compositions in 1964 and 1968 respectively. The comparison of job level compositions should be interpreted with care (see section 2.3).

The distribution of job families in 1956 and 1964 are exposed in table 3:4. The technical job families dominate. In 1956 13 % were employed in commercial work and 6 % in financial and office work. Each of the proportions had increased by 2 percentage units by 1964. Among the technical families the proportions of Managing and supervision of production and Design engineering decreased by 3 percentage units and 4 percentage units to 17 % and 26 % respectively by 1964, while the proportion of Research, experimental and development work increased by 2 percentage units to 16 %. The distribution for 1968 is given in table 3:5 without comparison, except for Research, experimental and development work which was in principle unaffected by the revision. Its proportion decreased to 13.5 % by 1968. Administrative work (0. in the new classification) has a higher than average proportion of employees with academic

Table 3:4. *Distribution by job families in 1956 and 1964; all educational qualifications*

Job family	1956	1964-1956	1964
1. Managing and supervision of production	0.201	–0.032	0.169
2. Research, experimental and development work	0.135	0.021	0.156
3. Design engineering, industrial and fashion design	0.296	–0.039	0.257
4. Other technical work	0.173	0.006	0.179
5. Humanistic and artistic work	–	0.001	0.001
6. Teaching	0.001	0.002	0.003
7. General services and health care	0.003	0.003	0.006
8. Commercial work	0.133	0.020	0.153
9. Financial and office work	0.057	0.019	0.076
All families	0.999	0.001	1.000
Total number of employees	27 449	17 139	44 588

Note: Jobs with less than 10 employees are excluded.

degrees. This category also includes legal work. Research, experimental and development work is more graduate engineer intensive than average, while production work and design are high-school engineer intensive. Besides administrative work, those with an economic education naturally work in commercial, financial and office jobs. The proportions of engineers are however relatively high in these job families also. The explanation is that commercial work often requires technical knowledge to understand and explain products to customers. These jobs also frequently involve a technical service function to the customers. A technical element in the job family Financial and office work is systems work, programming and computer operation. Of the 85 employees in General services and health care 63.5 % are physicians.

Some summary measures of changes in the educational composition have been calculated for each job family (table 3:6). The increase in the total number of employees in all families over the period 1956–1964 is recorded as 62 %. Only Managing and supervision of production and Design show a lesser increase. The changes in educational mix is largest in General services and health care, Commercial work and Financial and office work. The reason why the figure for General services and health care is so high is because in this family only Degrees

Table 3:5. *Educational composition by job families in 1968*

Education	Job family 0	1	2	3	4	5	6	7	8	9	All job families
Degree in engineering	0.193	0.114	0.284	0.088	0.057	–	–	–	0.067	0.038	0.109
Certificate in engineering I	0.182	0.341	0.342	0.369	0.335	0.393	0.581	–	0.302	0.090	0.317
Certificate in engineering II	0.218	0.529	0.323	0.543	0.599	0.607	0.419	0.365	0.411	0.114	0.454
Degree in business & economics	0.149	–	–	–	–	–	–	–			0.021
Certificate in commerce	0.099	–	–	–	0.003	–	–	–	0.178	0.587	0.081
Degree in social work	0.033	–	–	–	–	–	–	–	–	–	0.001
Degree in science	–	–	0.032	0.001	–	–	–	–	0.003	0.016	0.006
Degree in law or in the social sciences	0.125	–	–	–	–	–	–	–	0.003	0.007	0.003
Other academic degrees	–	0.016	0.018	–	0.006	–	–	0.635	0.003	–	0.008
All educational qualifications	0.999	1.000	0.999	1.001	1.000	1.000	1.000	1.000	1.001	0.000	1.000
Number of employees	1 097	10 215	7 304	13 113	7 866	61	129	85	9 873	4 208	53 951
Relative distribution of employees	0.020	0.189	0.135	0.243	0.146	0.001	0.002	0.002	0.183	0.078	0.999

Note 1: Employees employed by members of the Swedish Commercial Employers' Association, Central Group are covered by the survey in 1968 but not in 1956 and 1964. In 1968 this group totalled 1081 employees.

Note 2: Jobs with less than 10 employees are excluded.

Table 3:6. Q *and* S *measures of changes in educational composition in 1956–1964, by job families*

Job family★	$Q_{56,64}$	$S_{56,64}$	$\sqrt{\frac{n}{2}}\, S_{56,64}$	r
All families	1.624	0.014	0.030	1.000
1. Managing and supervision of production	1.366	0.013	0.019	0.530
2. Research, experimental and development work	1.870	0.017	0.026	0.607
3. Design engineering, industrial and fashion design	1.411	0.014	0.018	0.599
4. Other technical work	1.680	0.007	0.011	0.354
6. Teaching	3.389	0.029	0.029	0.630
7. General services and health care	2.761	0.179	0.334	0.205
8. Commercial work	1.878	0.029	0.051	0.952
9. Financial and office work	2.176	0.039	0.067	0.469

★ 5. Humanistic and artistic work is omitted.

Note: Jobs with less than 10 employees are excluded.

in law and social sciences and Other degrees were recorded in 1956, while all educational qualifications except degrees in engineering and science were recorded in 1964. A closer study of the distributions of the last two families reveals that in Commercial work the proportion of graduate engineers decreased from 12 to 9 % and that of high school engineers from 43 to 39 %, while employees with a certificate in commerce increased from 12 to 18 %, and in Financial and office work the proportion of graduate engineers increased from 0 to 1 %, that of high school engineers I from 5 to 8 % and those with a certificate in commerce from 58 to 64 %. The proportions of high school engineers II and graduate business economists decreased from 12 to 10 % and from 24 to 17 % respectively. In Commercial work the change in mix is very much the same as for all families, but in Financial and office work less so. As the correlation coefficient in table 3:6 indicates there are no families with changes opposed to the common change.

The distribution of job levels (table 3:7) shows that the middle levels are most frequent and that jobs at levels 7 and 8 are very rare. The changes in level composition revealed by the table are a decrease in the proportion of the top level jobs and an increase primarily at level 6. This change may be at least partly artificial and result from a downgrading of jobs during the years immediately after the introduction of the classification system (1956) and in connection with the revision in 1965/66. However, there may also be an effect due to increased output from schools and universities (see below), as initial jobs are usually at levels 5 and 6.

Table 3:8 shows that the academic proportions are high at the top levels. These groups both start at a higher level and are promoted to higher levels more frequently than non-academics. This does not mean, however, that non-academics do not reach the top level 2. In 1968 38 % of all educated employees at this level did not possess an academic degree.

Table 3:7. *Composition by job levels in 1956–1968; all educational qualifications*

Job level	1956	(1964–1956)	1964	(1968–1964)	1968	(1968–1956)
2	0.043	−0.012	0.031	−0.001	0.030	−0.013
3	0.132	−0.010	0.122	−0.004	0.118	−0.014
4	0.276	−0.002	0.274	−0.006	0.268	−0.008
5	0.379	0.008	0.387	−0.007	0.380	0.001
6	0.139	0.014	0.153	0.018	0.171	0.032
7	0.030	0.002	0.032	0.000	0.032	0.002
8	0.001	−0.001	0.000	0.001	0.001	0.000
Totals	1.000	−0.001	0.999	0.001	1.000	0.000
Total number of employees	27 449		44 588		53 951	

Note: See notes 1 and 2 table 3:5.

Table 3:8. *Educational composition by job levels in 1968*

Education	Job level 2	3	4	5	6	7	8	All levels
Degree in engineering	0.472	0.279	0.144	0.057	0.007	–	–	0.109
Certificate in engineering I	0.168	0.259	0.309	0.331	0.366	0.313	–	0.317
Certificate in engineering II	0.128	0.261	0.410	0.521	0.552	0.521	0.483	0.454
Degree in business & economics	0.121	0.054	0.026	0.010	–	–	–	0.021
Certificate in commerce	0.081	0.079	0.085	0.073	0.074	0.166	0.517	0.081
Degree in social work	–	0.002	0.002	–	–	–	–	0.001
Degree in science	0.012	0.017	0.010	0.004	–	–	–	0.006
Degree in law or in the social sciences	–	0.010	0.005	0.003	–	–	–	0.003
Other academic degrees	0.018	0.037	0.009	0.001	–	–	–	0.008
Totals	1.000	0.998	1.000	1.000	0.999	1.000	1.000	1.000
Total number of employees	1 634	6 375	14 473	20 525	9 201	1 714	29	53 951

Note: See notes 1 and 2 table 3:5.

The S measures in table 3:9 reveal some variability of changes in job mix between educational qualifications. The correlation coefficients also show a variability in direction. The decrease of the proportions at levels 2 and 3 holds for all educational qualifications. The proportions of almost all educational qualifications increase at level 5. The high S score for employees with a degree in law or in the social sciences is probably explained by the entry at levels 4 and 5 of relatively few employees with degrees in the social sciences. In 1956 this group was completely dominated by lawyers at level 3.

3.2.3 *Age composition*

The increased proportion of employees at job levels 5 and 6 is probably associated with a decrease in average age due to employment of young people. The overall age distribution is given in table 3:10 where it is also possible to study the changes in distribution over the period 1956–1968. The proportions of

Table 3:9. Q *and* S *measures of changes in job levels in 1956–1968, by educational qualification*

Education	$Q_{56,64}$	$Q_{64,68}$	$Q_{56,68}$	$S_{56,68}$	$\sqrt{\frac{n}{2}}S_{56,68}$	$r_{56,68}$
All educational groups*	1.623	1.218	1.964	0.014	0.027	1.000
Degree in engineering	1.466	1.106	1.621	0.045	0.072	0.277
Certificate in engineering I	1.674	1.231	2.061	0.035	0.061	0.906
Certificate in engineering II	1.526	1.207	1.841	0.012	0.023	0.506
Degree in business & economics	1.733	1.217	2.109	0.066	0.105	–0.035
Certificate in commerce	2.532	1.272	3.220	0.031	0.057	–0.165
Degree in science	3.000	1.386	4.159	0.181	0.256	0.294
Degree in law or in the social sciences	1.482	2.337	3.464	0.474	0.581	0.238
Other academic degrees	1.898	1.123	2.132	0.061	0.086	0.229

* Degree in social work is omitted. *Note:* See note 1 table 3:5.

Table 3:10. *Age distribution in 1956–1968; all educational qualifications*

Age interval	1956	(1964–1956)	1964	(1968–1964)	1968	(1968–1956)
–25	0.070	0.020	0.090	0.005	0.095	0.025
26–29	0.144	–0.001	0.143	0.010	0.153	0.009
30–34	0.244	–0.072	0.172	0.001	0.173	–0.071
35–44	0.350	–0.002	0.348	–0.056	0.292	–0.058
45–59	0.162	0.057	0.219	0.039	0.258	0.096
60–	0.030	–0.002	0.028	0.000	0.028	–0.002
Totals	1.000	0.000	1.000	–0.001	0.999	–0.001
Number of employees	28 693		45 737		55 728	

Note: See note 1 table 3:5.

employees in the group 30–44 years have decreased while both younger and older employees become relatively more numerous. The same change has taken place for almost all educational groups. For high school engineers II the increase in proportion in the group 45–60 years is higher than for any other education. There is also some relative increase in engineers 60 years and older. Those who have a certificate in commerce deviate a little from the other educational groups because the proportions of those who are less than 34 years old have decreased while the other proportions have all increased, (see the correlation coefficients in table 3:11).

Table 3:11. *Changes in age mix by education in 1956–1968*

Education	$S_{56,64}$	$S_{64,68}$	$S_{56,68}$	$\sqrt{\frac{n}{2}}S_{56,68}$	$r_{56,64}$	$r_{64,68}$	$r_{56,68}$
All educational qualifications	0.038	0.029	0.055	0.095	1.000	1.000	1.000
Degree in engineering	0.015	0.014	0.019	0.033	0.539	0.888	0.864
Certificate in engineering I	0.025	0.020	0.032	0.055	0.835	0.978	0.890
Certificate in engineering II	0.066	0.045	0.094	0.163	0.984	0.997	0.992
Degree in business & economics	0.032	0.024	0.038	0.066	0.827	0.761	0.801
Certificate in commerce	0.033	0.018	0.043	0.074	0.004	0.206	–0.016
Other unspecified educational qualifications	0.022	0.046	0.064	0.111	0.429	0.378	0.230

Note: See note 1 table 3:5.

3.3 SUPPLY OF LABOUR FROM UNIVERSITIES AND SCHOOLS

The supply of young labour from high schools and universities has increased very much during the sixties. Of the educational groups which to any extent enter industry, the increase has been most pronounced for the non-academic educational qualifications. This is clearly seen in figure 3:3 and table 3:12. The intake to the colleges of technology and economics , although showing a slow expansion, has been limited during the whole period. The so called educational explosion has at the academic level mainly taken place in the social and humanistic sciences, which have not so far become an important source of recruitment for Swedish industry. The explosion reached its peak after 1968, the last year covered by this study.

Although figure 3:3 and table 3:12 only give the number of degrees and certificates awarded, the educational explosion is not only an increase relative to previous output from schools and universities but also an increase to the population in the relevant age intervals. For instance, between 1964 and 1968 the population in the age interval 15–24 years increased by 5.6 % while the number of certificates awarded annually in engineering I increased by 88.7 %, certificates in commerce by 264.3 % and degrees in engineering by 46 %.

Table 3:12. *Number of degrees and certificates awarded in 1954–1968*

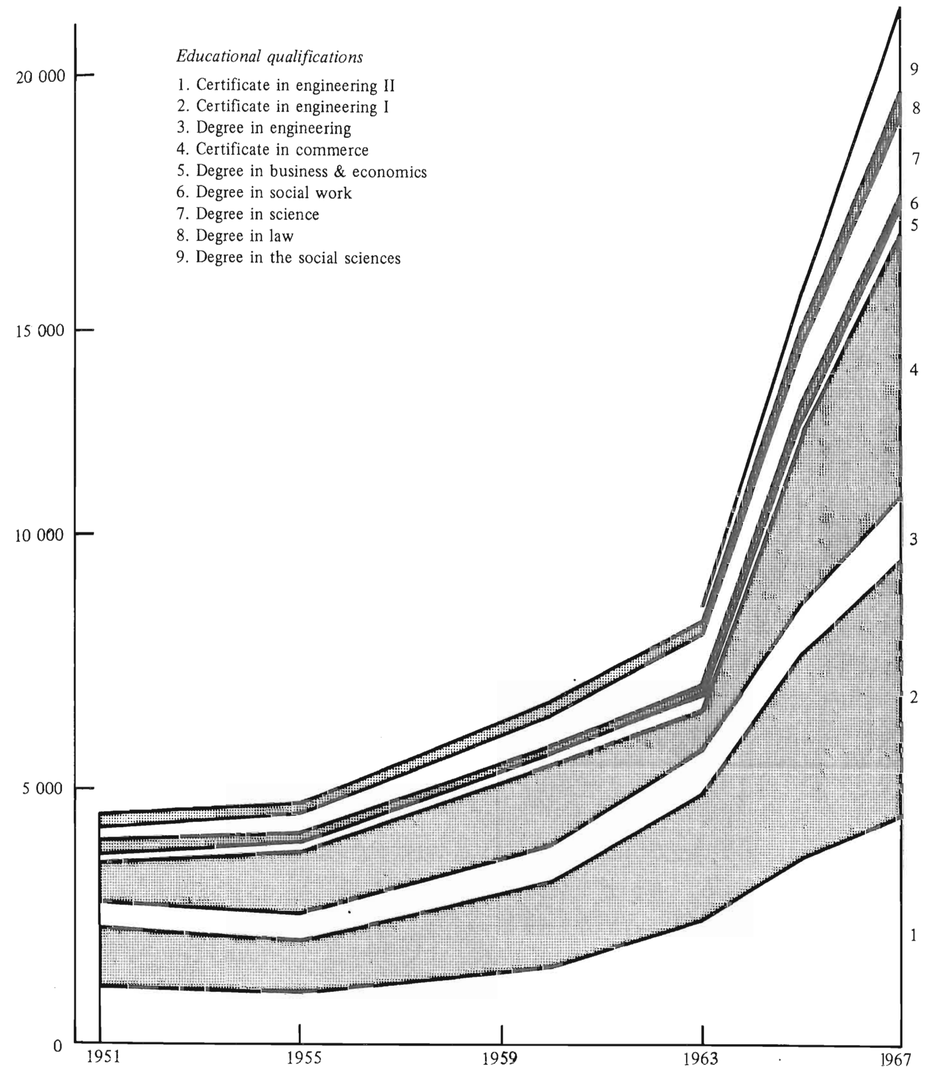
Education	1954–1958	1959–1963	1964–1968	1954–1968
Degree in engineering (Civ.ing., Tekn.lic., Tekn.dr)	2 832	3 978	6 031	12 841
Certificate in engineering I (Läroverksingenjör)	5 453	9 352	23 184	37 989
Certificate in engineering II (Institutsingenjör)★	5 747	9 555	15 644★★	30 946★★
Degree in business & economics (Civ.ekon., Ekon.lic., Ekon.dr)	988	1 268	2 244	4 500
Certificate in commerce (Handelsgymnasie-examen)	5 867	7 208	21 395	34 470

★ Includes only schools supported by the Government. In 1961 private schools passed 540 students and in 1965 a maximum of 1 128.

★★ Figures only available to 1967.

Source: Statistical year book for Sweden and unpublished data from the Central Bureau of Statistics.

Figure 3:3. *Degrees and certificates awarded in 1951–1967*
(Cumulative scale)



Source: See table 3:12.

Note: 1. only includes schools supported by the Government.
9. was first recorded as a separate group in 1964.

Finally a word of caution. The number of degrees and certificates awarded does not correspond to the same number of individuals, since an individual may obtain more than one degree or certificate. For instance, some of the students who obtain a certificate in engineering I continue to a college of technology to work for a degree. However, this can in no way disturb the observed general tendency.

3.4 SOME EVIDENCE ON EMPLOYMENT, LEAVING AND PROMOTION RATES

A breakdown of the salary statistics into three groups, Leavers, Pairs and Beginners, was described in chapter 2. It provides some insight into labour mobility. If mobility is low one would expect a rigid job mix, industry mix, and so forth as well as a rigid salary structure, but if the mobility is high a rigid mix and perhaps also a rigid salary structure are more noteworthy, (see section 5.2.7).

The data used for this study are from 1963 and 1964. As they were not originally produced for a study of labour mobility, they leave shortages and unanswered questions, for instance with respect to job changes and moves between industries. As a reminder: those who are not registered with the same employer in 1964 as in 1963 are Leavers (L63), those who are registered with the same employer in both 1963 and 1964 are Pairs (P63 and P64), and those who are not registered with the same employer in 1963 as in 1964 are Beginners (B64). The categories Leavers and Beginners partly cover the same individuals. Those who leave a SAF employer but are not employed by another SAF employer and those who obtain promotion to the management level (whether or not with the same employer) are no longer registered in the statistics and are therefore only included among the Leavers. In the same way, salary earners who are newly employed and do not come from a SAF member and wage earners who obtain promotion to the job of work supervisor are included only among the Beginners.

In table 3:13 it is possible to investigate how leaving rate, employment rate and promotion rate are associated with a number of variables. Unfortunately, data are not available to permit cross classifications. Interactions between variables cannot be investigated. The data used are the same as in model I, chapter 5.2 (p.169) and they cover engineers in technical job families. (For a more detailed description see Klevmarken [1968a], part II.)

Between August 15th, 1963 and August 15th, 1964, 15 % of the engineers employed by SAF members covered by the data left. 18 % of the engineers registered on August 15th, 1964 were employed during the same period. Although methods and definitions are not exactly the same in the investigation made by the Swedish Labour Board in 1968 (Faragó[1968]), the results obtained still provide an interesting comparison. Of a total of 3 200 000 employees in Sweden, 566 000 i.e. 17.7 % had moved from one employer to another during the period May 1st, 1965 – April 30th, 1966. This figure is obtained for a business cycle peak and it is probably a little higher than it would have been for the period August 15th, 1963 – August 15th, 1964, but disregarding this difference the turnover rate for the salaried employees studied seems to be of the same magnitude as the average turnover rate for the whole Swedish labour market.

As other investigations have shown, mobility is higher among young employees than among old ones. The leaving rate is 25 % for the youngest and then decreases to a minimum of 9 % in the age interval 45–59 years. Because of retirement it is high again, 24 %, for those who are 60 years or older. Except for this age interval the employment rate shows a similar age pattern. 55 % of the youngest engineers were employed during the previous year, while only 5 % of the oldest ones are newly employed.

Both leaving rate and employment rate are the lowest in cost of living area 3 and the highest in area 5. The differences are however small. The differences between the three educational qualifications are also small. Graduate engineers show slightly lower rates than high school engineers II, while high school engineers I show some higher rates. These differences do not necessarily have anything to do with education, but can possibly be explained by differences in age distribution. High school engineers I have the lowest average age while graduate engineers and high school engineers II have about the same average. Table 3:14 shows leaving rates★ and employment rates for a larger sample of educational qualifications and jobs. These data show that employees with an education in economics have higher rates than the engineers. Their average age is, however, lower than the average of graduate engineers and high school engineers II but not as low as that of high school engineers I. The leaving rates and employment rates for the small educational groups are based on a very small number of observations and the observed deviations from the average should not be given too much importance.

A comparison of the mobility by industry (table 3:13) reveals more variability between industries than between educational groups. There is high mobility in Building and construction and in Repair works, while Mining, Iron and steelworks, Manufacture of hardware, Food manufacturing industries and Beverage and tobacco industries have a relatively low engineer turnover. The differences observed are difficult to interpret for the same reason as above. We do not know anything about differences in age and job distribution. The high figure for Building and construction may, however, be a result of the favourable business conditions for this industry in 1963/64.

★ Leaving rates are calculated with Pairs 1964 + Leavers 1963 as the base instead of Pairs 1963 + Leavers 1963.

The turnover of labour is higher at low job levels than at higher level jobs. As young employees are most frequent at low job levels, it is difficult to separate this effect from the age effect. A comparison of employment and leaving rates and the difference between Pairs 1963 and Pairs 1964 may also be of some interest. For all job families enumerated in table 3:13 there are 1098 engineers at level 6 who left, 599 who obtained promotion within the same company and 1741 who were employed. At level 3 we find that 293 left, 207 obtained promotion within the company to this job level and 242 were employed. At low job levels labour is usually recruited outside the company★ while promotion within the company accounts for a large part of the recruitment at higher job levels. But even at these levels, employment from outside amounts to about 50 % of the supply.

A comparison between job families is not without difficulty because some job levels are missing in many families. The comparison also varies somewhat from job level to job level. The sum up, for the technical job families studied it may be stated that the turnover of engineers was about the same for all job families except for Productivity engineering (400) where it was a little higher.

A similar comparison between job families has been made using the same data as those in table 3:14. The turnover rate has this time been defined as the minimum of employment rate and leaving rate. Since job levels are not the same in each job family, it is not meaningful to calculate an average of the turnover rates and to compare the results. For each job level the families have instead been ranked by turnover rate. The result is given in table 3:15. There is no unique ranking of the job families. As a tentative conclusion we may state that the turnover is relatively high in Work supervision, Personnel work and Commercial work, while it is low in Mathematical work and Laboratory work.

★ Note that »Beginners» also include wage earners promoted to work supervisors within the same company.

Table 3:13. *Leaving, employment and promotion rates for engineers in technical work*

Variable	Rates (%) L63/(L63+P63)	B64/(B64+P64)
Age		
–25	24.7	54.5
26–29	20.5	27.9
30–34	16.4	16.6
35–44	11.9	11.1
45–59	8.9	7.4
60–	23.9	5.3
All age intervals	14.9	18.2
Cost of living area		
3	13.6	17.1
4	14.3	17.9
5	18.5	21.2
All areas	14.9	18.2
Education		
Degree in engineering	13.1	14.7
Certificate in engineering I	15.4	19.4
Certificate in engineering II	15.2	18.5
All educational qualifications	14.9	18.2
Industry		
Mining	13.3	9.5
Metal and engineering industry		
Iron & steel works, metal plants	11.0	14.2
Manufacture of hardware	11.1	14.4
Engineering works	16.6	17.9
Repair works	20.4	24.7
Shipyards	15.2	16.0
Manufacture of electrical equipment	12.9	19.1
Other metal industry	12.2	16.1
Quarrying; stone, clay and glass products	11.4	20.7
Wood industry	14.1	21.5
Manufacture of pulp, paper and paper products	13.6	13.0
Printing and allied industries	15.2	17.6
Food manufacturing industries	10.4	12.4
Beverage and tobacco industries	10.9	12.3
Textile industry	13.8	12.0
Leather, furs and rubber industries	9.2	21.2
Chemical industry	14.4	18.7
Building and construction	19.7	24.7
All industries	14.9	18.2

Table 3:13, cont.

Job

Family		Level 2	3	4	5	6	7
Supervision of production (110)	L63/(L63+P63)	12.9	11.1	10.7	13.8		
	B64/(B64+P64)	10.1	9.7	9.3	17.6		
Work supervision, general (120)	L63/(L63+P63)			13.5	18.8		
	B64/(B64+P64)			16.0	17.1		
Work supervision, building and construction work (130)	L63/(L63+P63)				16.0	30.2	38.7
	B64/(B64+P64)				18.7	38.6	55.6
Mathematical work (200)	L63/(L63+P63)	13.3	5.3	8.1	13.8	23.2	
	B64/(B64+P64)	6.3	3.8	8.3	13.7	39.8	
Laboratory work (210)	L63/(L63+P63)	14.2	7.7	10.0	12.8	19.9	24.7
	B64/(B64+P64)	4.2	6.4	12.1	15.2	3.4	4.5
Mechanical and electrical design (310)	L63/(L63+P63)	14.7	9.7	9.4	14.7	20.4	26.6
	B64/(B64+P64)	11.0	7.6	7.6	15.7	31.9	52.5
Productivity engineering (400)	L63/(L63+P63)	20.9	12.5	12.8	20.7	24.2	
	B64/(B64+P64)	7.0	6.9	14.2	27.1	43.3	
Technical instruction and technical service (470)	L63/(L63+P63)		9.9	10.8	15.2	20.7	
	B64/(B64+P64)		8.0	12.7	22.1	43.3	
Other internal technical work (490)	L63/(L63+P63)	16.9	7.8	9.6	15.3		
	B64/(B64+P64)	5.1	4.0	6.5	14.8		
All families	L63/(L63+P63)	14.2	9.8	10.6	15.0	22.2	28.3
	B64/(B64+P64)	8.8	7.7	9.7	17.3	34.9	51.6

L63 = Leavers 1963
P63 = Pairs 1963/64, observed 1963
P64 = Pairs 1963/64, observed 1964
B64 = Beginners 1964

The total number of observations are
L63: 4 124
P63: 23 500
P64: 23 283
B64: 5 188

Note: All Pairs 1964 differ from all Pairs 1963 because 217 engineers have moved to jobs not covered by the sample. Most of the 217 engineers were high school engineers employed in Engineering works and in Manufacture of electrical equipment in cost of living area 5.

Table 3:14. *Leaving rates and employment rates by education in 1963 and 1964*

Education	Number of employees			Rates (%)	
	L63	P64	B64	L63/(L63+P64)	B64/(B64+P64)
Degree in engineering	689	4 414	802	13.5	15.4
Certificate in engineering I	1 918	10 780	2 535	15.1	19.0
Certificate in engineering II	2 867	15 543	3 535	15.6	18.5
Degree in business & economics	181	744	183	19.6	19.7
Certificate in commerce	495	2 282	579	17.8	20.2
Degree in social work	11	34	17	24.4	33.3
Degree in science	28	242	57	10.4	19.1
Degree in law or in the social sciences	26	67	10	28.0	13.0
Other academic degrees	37	220	54	14.4	19.7
All educational qualifications	6 252	34 326	7 772	15.4	18.5

Note: The table is based on the following job families:

110	200	310	400	450	780	810	860	900	960
120	210	315	405	470		820	870	910	
130		330	410	480		840		940	
			440	490		850		945	

and on job levels 2–8 when applicable.

Table 3:15. *Ranking of job families by turnover rate*
Least turnover = rank 1

Family	Level					
	2	3	4	5	6	7
Supervision of production (110)	4	4	2	3		
Work supervision (120, 130)			7	6	6	4
Mathematical and laboratory work (200, 210)	2	1	3	2	4	1
Design engineering (310, 315, 330)	5	2	1	4	3	2
Other technical work (400, 405, 410, 440, 450, 470, 480, 490)	1	3	4	5	5	
Personnel work (780)	7	7	5	1		
Commercial work (810, 820, 840, 850, 860, 870)	6	6	8	7	1	
Financial and office work (900, 910, 940, 945, 960)	3	5	6	8	2	3

Note: Turnover rate is defined as the smallest of the ratios L63/(L63+P64) and B64/(B64+P64).

CHAPTER 4

AGE–EARNINGS PROFILES BY EDUCATION

4.1 INTRODUCTION AND SUMMARY

The purpose of this chapter is to develop a statistical framework for investigation of age-earnings profiles. A model is derived from a simple representation of individual earnings as the sum of initial earnings and accumulated yearly increases. Initial earnings are also assumed to change from one year to another. By the application of different constraints to the changes in earnings various models of age-earnings profiles can be obtained. In this chapter a fairly simple model is applied to salary data from Swedish industry. The estimates obtained are used to compare salary differences between employees with different educational qualifications by means of estimated profiles, lifetime salaries and relative rates of return for education.

A similar work was done by Fase in his interesting study from 1969. He assumed that the distribution of individual incomes at each age can be described by a log-normal function which varies systematically with age. An individual is supposed to start his career at the age of s years. The initial salary of an individual s years old is drawn from a lognormal distribution with parameters which in principle depend on s. A systematic effect and a random disturbance contribute to the annual change in *log* income. Income is supposed to increase up to a certain age and then decrease by the *same* rate. The annual rate of increase depends linearly on age, being the largest at s and declining with age. The random disturbance is obtained from a normal distribution. Because the random components of the model are assumed to follow a specific family of distributions, Fase is able to use maximum likelihood methods to estimate the unknown parameters. These are the change in rate of increase of income (i.e. the slope in the linear function of age), the age when income reaches a maximum, the expected initial log income, the variance of initial log incomes and the variance of the random disturbance to income increases.

Although there are great similarities between Fase's study and the present study there are also distinct differences. Fase does not distinguish between cross section profiles and cohort profiles. He applies his model to cross sections. In this study the distinction between the two profiles and also the relation between them are essential. The model developed below is applied to *pooled* cross sections covering between ten and twenty years. Fase's assumption that the rate of increase in salary is a decreasing linear function of age is an approximation too crude for some applications. Our model is more flexible on this point. Furthermore, a distinction is made in this study between physical age and the time spent in the labour market, referred to as active age. Fase's model only has the physical age dimension, and in his derivation of an estimation method he assumes that s, i.e. the age at which an individual joins the labour market, is constant. This assumption is rarely satisfied in applications. Fase's explicit assumption about a normal distribution formally allows rather precise statements about estimates and about quantities derived from the estimates as for instance lifetime salaries. Similar precise statements could be made with the model in the present study if the same assumptions about normality were adopted. The realism of this is not commonly accepted, however, and as it is not necessary to specify a particular distribution to obtain estimates the general strategy is not to do so. (There is one exception in section 4.3.5.) For the same reason least squares estimation is used instead of maximum likelihood estimation.

4.2 A FORMAL REPRESENTATION OF AN INDIVIDUAL EARNINGS PATH

Consider a group of n persons at time T. For the moment only the following characteristic of the group will be defined. There are n_0 individuals who obtain employment at time 0, n_1 at time 1, and so forth.

$$n = \sum_{t=0}^{T} n_t ; \tag{4:1}$$

The first subgroup has, with possible interruptions, been on the labour market for T time periods, the second group for T−1 periods and so forth. The number of years which have elapsed since a person first obtained employment is called his active age below. For the moment we will only assume that it is possible to determine the active age for each individual, without specifying how this could be done.

The notation for time is chosen by convenience. Time 0 is just a reference point which may be substituted for any convenient calendar time.

The initial salary for one of those who started work at t=0 can be defined as

$$\exp[\alpha + e_{0i}]; \quad i=1,\ldots,n_0 \tag{4:2}$$

when e_{0i} is defined* in such a way that

$$\sum_{i=1}^{n_0} e_{0i} = 0; \tag{4:3}$$

* The exponential function is in the following always denoted exp, while e means a residual.

This definition implies that $\exp[\alpha]$ can be interpreted as the (geometrical) average* initial salary of the n_0 persons and $\exp[e_{0i}]$ as the deviation from the average of individual i.

This average initial salary changes by

$$(\exp[\beta_t] - 1)100; \quad t=1,...,T \tag{4:4}$$

per cent from time t–1 to time t.

The average initial salary of a subgroup of individuals, who started work at time b, can be written as

$$\exp[\alpha + \sum_{t=1}^{b} \beta_t]; \quad b=1,...,T \tag{4:5}$$

and the initial salary of an individual belonging to the group is

$$\exp[\alpha + \sum_{t=1}^{b} \beta_t + e_{bi}]; \quad b=1,...,T. \quad i=1,...,n_b \tag{4:6}$$

$\exp[e_{bi}]$ is individual i's deviation from the average. By definition these deviations obey the constraint

$$\sum_{i=1}^{n_b} e_{bi} = 0; \quad b=1,...,T \tag{4:7}$$

The average salary of those who started work at time b changes from time period to time period. The average change from t–1 to t is

$$(\exp[\gamma_{tb}] - 1)100; \quad b=1,...,T. \quad t=b+1,...,T \tag{4:8}$$

per cent, and the change in the i-th individual is defined as

$$(\exp[\gamma_{tb} + u_{tbi}] - 1)100; \quad b=1,...,T. \quad t=b+1,...,T. \quad i=1,...,n_b \tag{4:9}$$

per cent. Again, by definition, the sum of all individual deviations from the average equals zero.

$$\sum_{i=1}^{n_b} u_{tbi} = 0; \quad b=1,...,T. \quad t=b+1,...,T \tag{4:10}$$

The average salary at time T for those who obtained employment at time b is then

$$L_{Tb.} = \exp[\alpha + \sum_{t=1}^{b} \beta_t + \sum_{t=b+1}^{T} \gamma_{tb}]; \quad b=1,...,T \tag{4:11}$$

* 'Average' will in the following stand for geometrical average when nothing else is said or when another interpretation is not obvious from the context.

The full stop in the expression $L_{Tb.}$ indicates an average. The corresponding individual salary is

$$L_{Tbi} = L_{Tb.} \exp[e_{bi} + \sum_{t=b+1}^{T} u_{tbi}]; \quad b=1,\ldots,T. \quad i=1,\ldots,n_b \tag{4:12}$$

This representation of the individual salary growth is summarized in table 4:1. It should perhaps be emphasized that the scheme presented is not a *model* of individual salary growth, it is only a formal framework which can be used to represent any salary path for any group of individuals. In the following it will be used as a starting point for model building, and this is of course the purpose of presenting this particular scheme.

Before work on the model begins in section 4.3 it may be useful to carry out some preliminary exercises to examine some properties of this formal representation.

A row in the table represents a cohort profile while the last column represents a complete cross section profile. A typical element in an average cohort profile is thus given by expression (4:11) for a given $b = b_o$ and any T in an interval determined by b_o and the maximum active age. Suppose for the sake of simplicity that there is a maximum active age independent of time, say $(T-b)_{max}$. The interval is then $b_o \leqslant T \leqslant b_o + (T-b)_{max}$. The average cohort profile is

$$L_{Tb_o.} = \exp[\alpha + \sum_{t=1}^{b_o} \beta_t + \sum_{t=b_o+1}^{T} \gamma_{tb_o}]; \tag{4:13}$$

Individual cohort profiles are obtained analogously. A typical element in an average cross section profile is obtained from the same expression (4:11) for a given $T = T_o$ and any b in the interval $T_o-(T-b)_{max} \leqslant b \leqslant T_o$.

$$L_{T_o b.} = \exp[\alpha + \sum_{t=1}^{b} \beta_t + \sum_{t=b+1}^{T_o} \gamma_{tb}] = \exp[\alpha + \sum_{t=1}^{T_o} \beta_t + \sum_{t=b+1}^{T_o} (\gamma_{tb} - \beta_t)]; \tag{4:14}$$

As cross section profiles involve comparisons between individuals, it is not meaningful to define an individual cross section profile.

The two expressions (4:13) and (4:14) reveal that the shape of a cohort profile is determined by the increments γ_{tb} while the shape of a cross section profile is determined by the differences $(\gamma_{tb}-\beta_t)$. Provided that γ_{tb} and β_t are greater than zero the slope of the cohort profile will thus be steeper than the slope of a cross section profile (see figure 1:1) and if the difference $(\gamma_{tb}-\beta_t)$ is negative the cross section profile will decrease with increasing active age (T_o-b).

Table 4:1. *Earnings paths for individuals with different periods of entry*

Period of entry	Time 0	1	2	. . .	T
0	$\exp[\alpha+e_{0i}]$;	$\exp[\alpha+e_{0,i}+\gamma_{10}+u_{10i}]$;	$\exp[\alpha+e_{0,i}+\sum_{t=1}^{2}(\gamma_{t0}+u_{t0i})]$;	. . .	$\exp[\alpha+e_{0i}+\sum_{t=1}^{\Gamma}(\gamma_{t0}+u_{t0i})]$;
1		$\exp[\alpha+\beta_1+e_{1i}]$;	$\exp[\alpha+\beta_1+e_{1i}+\gamma_{21}+u_{21i}]$;	. . .	$\exp[\alpha+\beta_1+e_{1i}+\sum_{t=2}^{T}(\gamma_{t1}+u_{t1i})]$;
2			$\exp[\alpha+\sum_{t=1}^{2}\beta_t+e_{2i}]$;	. . .	$\exp[\alpha+\sum_{t=1}^{2}\beta_t+e_{2i}+\sum_{t=3}^{T}(\gamma_{t2}+u_{t2i})]$;
.				. . . .	. . . .
T−1					$\exp[\alpha+\sum_{t=1}^{T-1}\beta_t+e_{(T-1)i}+\gamma_{T(T-1)}+u_{T(T-1)}$
T					$\exp[\alpha+\sum_{t=1}^{T}\beta_t+e_{Tti}]$;

Both profiles can be used to calculate lifetime salaries. The calculations can be made in many ways, but the following definitions are chosen. From the cohort profile (4:13) a sum of average salaries is obtained

$$S_{co} = \sum_{T=b_o}^{b_o+(T-b)_{max}} L_{Tb_o\cdot} \; ; \qquad (4{:}15)$$

and similar from the cross section profile (4:14)

$$S_{cr} = \sum_{b=T_o-(T-b)_{max}}^{T_o} L_{T_ob\cdot} \; ; \qquad (4{:}16)$$

An individual lifetime salary can be defined by analogy to (4:15), while there is no individual cross section lifetime salary as no individual cross section profile is defined.

The formal representation can be applied to earnings data in both current and constant prices. To see this, assume that the percentage change in the average price level between t−1 and t, measured by some convenient index is

$$(\exp[\pi_t] - 1)100; \quad t=1,\ldots,T \qquad (4{:}17)$$

If the deflated salary is denoted L^R and nominal salary L^N, we obtain

$$L^R_{Tb\cdot} = L^N_{Tb\cdot}\exp[-\sum_{t=1}^{T}\pi_t] = \exp[\alpha + \sum_{t=1}^{b}(\beta_t - \pi_t) + \sum_{t=b+1}^{T}(\gamma_{tb} - \pi_t)]; \quad (4{:}18)$$

In (4:18) time period 0 is the base. The average *real* increases in initial salary are thus $(\beta_t - \pi_t)$ and the average *real* increases for a cohort b are $(\gamma_{tb} - \pi_t)$. For a given b the expression (4:18) determines a cohort profile in constant prices. The age-dependent increments in an average cross section profile are of course independent of π_t while the starting point of the profile $L_{T_oT_o}$ is not.

$$L^R_{T_ob\cdot} = \exp[\alpha + \sum_{t=1}^{T_o}(\beta_t - \pi_t) + \sum_{t=b+1}^{T_o}(\gamma_{tb} - \beta_t)] \; ; \qquad (4{:}19)$$

The exercise above shows how average increments obtained from data in current prices can be transformed to increments from data in constant prices. We will return to this topic.

4.3 MODELS OF AGE–EARNINGS PROFILES

The formal representation will now be used for model building. More or less restrictive assumptions are imposed on the average increments β_t and γ_{tb} and the models are built up directly for application to average earnings (salaries) and not to individual earnings (salaries). The main reason for this approach is the difficulty of working with the large amount of individual data required. The assumptions used which are sometimes rather crude may also be more justified for averages than for individuals. The general strategy is firstly to discuss the non-stochastic properties of all model variants and then to discuss the stochastic properties. To introduce some general ideas a »prototype» model is first presented. It is then extended to a more general model and different variants of this model are formulated. As the model is constructed for application to average salaries the stochastic component of the model is also related to an average salary. In specifying the properties of the stochastic component we regard it as a sample average of individual deviations from the non-stochastic part of the model.

The time period used is a year. This is necessary because the data sources available only provide yearly observations, but it is also a natural periodization because negotiations and salary revisions are usually carried out on an annual (or multi-annual) basis.

4.3.1 *A prototype model with salary increases dependent on active age*

The increments β_t of the average initial salary are now assumed constant except for a random disturbance. The average initial logarithmic salary $\ln L_{bb.}$ is treated as a stochastic variable

$$\ln L_{bb.} = \alpha + \beta b + \epsilon_b \; ; \quad b=0,\ldots, \tag{4:20}$$

where α and β are constants and ϵ_b a stochastic variable with expectation zero.

In section 4.2 the sub-indexes t and b were attached to the average increments γ to show that they depend both on calendar time and on cohort. The dependence on calendar time should not be taken to mean that γ_{tb} follows a simple trend, but rather that γ_{tb} depends on other factors which have a unique effect on the salary changes each year. Examples are effects due to changes of supply and demand and to negotiations. These effects will be treated more extensively later on.

It is commonly accepted that salary increases are obtained as a result of the experience gained. Investment in training (on-the-job training) increases the marginal productivity and also earnings of an employee. But training takes time and therefore it is natural to assume that the increases γ_{tb} are a function of active age (t–b).

Although investment in training may take place over the whole range of active age it is mainly done immediately after entry into the labour market. During the training period marginal productivity is low and increases in earnings are thus rather low during the very first years on the labour market but then become higher as experience is gained.★ When investment later decreases, the increase in marginal productivity and earnings diminishes. The earnings profile would assume the general shape indicated by the curve TT in figure 4:1 which is a reproduction of figure 1 in Becker [1962]. UU represents the profile with no training. (T'T' is a more extreme case than TT with a training period limited to a few years.)

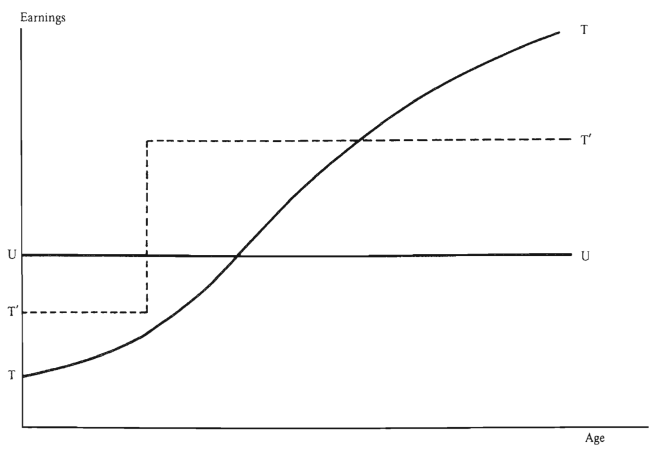
Differences in school education also have effects on the age-earnings profile, because more on-the-job training is usually given to those who have more schooling. When studying educational groups, earnings profiles of high school educated employees will then be flatter than profiles of employees with a university degree. (Mincer [1958] p. 292 has empirically found ... »that earnings are not only higher but also increase more rapidly with age (or decline more slowly after the peak of earnings is reached) in the more highly trained groups than in the less trained ones.»)

For the moment γ_{tb} is assumed to vary systematically with active age only. (A more complex explanation of γ_{tb} is introduced later.)

$$\gamma_{tb} = F(t-b) + \vartheta_{tb}; \qquad (4:21)$$

$$E(\vartheta_{tb}) = 0; \qquad (4:22)$$

Figure 4:1. *Age-earnings profiles*



★This may not be true if only specific training and no general training is given (see Becker [1962]).

Although we have an idea about the shape of the function F it may be difficult to find a functional form with parameters which can easily be interpreted and also give a good fit.★ For this reason the range is divided into c intervals with the limits $A_0, A_1, \ldots, A_c$ and the function F is approximated by a function which is constant within each interval. The constant values of the intervals of this function, γ_i, i=1,...,c, can be estimated from the data. With these assumptions and with assumptions (4:20) and (4:21), (4:11) can be rewritten as a polygon

$$L_{Tb.} = \exp[\alpha + \beta b + \sum_{i=1}^{c} \gamma_i D_i' + \epsilon_b + \sum_{t=b+1}^{T} \vartheta_{tb}]; \qquad (4:23)$$

where D_i' is the time spent in active age class i or★★

$$D_i' = \begin{cases} 0 & T-b < A_{i-1} \\ (T-b)-A_{i-1} & \text{if } A_{i-1} \leqslant T-b \leqslant A; \\ A_i - A_{i-1} & A_i < T-b \end{cases} \qquad (4:24)$$

and where

$$E(\epsilon_b + \sum_{t=b+1}^{T} \vartheta_{tb}) = 0; \qquad (4:25)$$

Table 4:2 may help to explain the notation more fully. It is similar to table 4:1 except that individual as well as stochastic components are omitted. The table is drawn up under the assumption that $A_0 = 0$ and $A_1 = 2$. Also in this table a cohort profile is obtained from a row, while a cross section profile is obtained from a column.

For a given b, (4:23) is the expression for an average cohort profile. From the definition of D_i' it follows that

$$b = T - \sum_{i=1}^{c} D_i' ; \qquad (4:26)$$

For a certain $T = T_0$ substitution of (4:26) into (4:23) gives the expression for a cross section profile

$$L_{T_0 b.} = \exp[\alpha + \beta T_0 + \sum_{i=1}^{c} (\gamma_i - \beta) D_i' + \epsilon_b + \sum_{t=b+1}^{T_0} \vartheta_{tb}] ; \qquad (4:27)$$

★ Fase [1969] used a linear function with a negative slope (see for instance Fase [1969] figure 4, p. 15) and the Swedish Employers' Confederation uses the function $L_{Tb} = \exp[a_0 + a_1 \frac{1}{T-b} + \ldots + a_n \frac{1}{(T-b)^n}]$; (SAF [1969]).

★★ For later use the notation D_i' is preferred to D_i.

Table 4:2. *Expected logarithmic cohort and cross section profiles generated by the prototype model* $[E(\ln L_{Tb.})]$

Year of entry (b)	Time (t) 0	1	2	3	4 . . .	T
0	α	$\alpha+\gamma_1$	$\alpha+2\gamma_1$	$\alpha+2\gamma_1+\gamma_2$	$\alpha+2\gamma_1+2\gamma_2$	$\alpha+\sum_{i=1}^{c}\gamma_i D_i'$
1		$\alpha+\beta$	$\alpha+\beta+\gamma_1$	$\alpha+\beta+2\gamma_1$	$\alpha+\beta+2\gamma_1+\gamma_2$	$\alpha+\beta+\sum_{i=1}^{c}\gamma_i D_i'$
2			$\alpha+2\beta$	$\alpha+2\beta+\gamma_1$	$\alpha+2\beta+2\gamma_1$	$\alpha+2\beta+\sum_{i=1}^{c}\gamma_i D_i'$
3						.
.						.
.						.
.						.
.						
T						$\alpha+T\beta$

4.3.2 *A model with salary increases dependent on active and physical age*

If we keep active age constant, it is very likely that the remaining variability can partly be explained by physical age. The physical and mental ability of a young employee are usually higher than those of an old employee, and this should have an influence on marginal productivity and earnings in addition to experience. The rate of increase in earnings γ_{tb} should therefore also be a function, probably a decreasing one, of physical age. There are also other reasons which suggest that physical age is a strategic variable. The salary statistics produced by the Swedish Employers' Confederation (SAF) play an instrumental role in the salary setting process as described in section 4.2. As this data source gives information about average salaries in different age groups it strengthens the common practice of setting salaries according to a person's age in addition to other criteria. In conclusion the rate of increase in earnings should vary with calendar time, active age and physical age.

The introduction of physical age into the model demands a small change in notation. Those who were born in year f and started work in year b received the average initial salary $L_{bfb.}$ and they will on average earn $L_{Tfb.}$ in year T. The average logarithmic initial salary deviates from its expected value by ϵ_{fb} which is a stochastic variable with zero expectation. The rate of increase between t−1 and t in the average salary is γ_{tfb}. Our new assumptions about the rate of increase in earnings can formally be written

$$\gamma_{tfb} = G\,[(t-b), (t\text{-}f)\,] + \vartheta_{tfb}\ ; \tag{4:28}$$

$$E(\vartheta_{tfb}) = 0; \tag{4:29}$$

There is no causal dependence between the rate of increase in earnings and active and physical age, but rather the ability to do a job, negotiation practice and other factors associated both with active and physical age. A specification of the function G would properly require an investigation of these relationships, but the approach taken here is to postulate that the active age and the physical age effects are additive and, as before, to approximate G by a step function. The limits of the physical age intervals are the same as those of the active age except for a suitably chosen constant: $A_0 + C,\ A_1 + C, ..., A_c + C$. There is thus a unique correspondence between an active age interval and a physical age interval. In analogy with D_i' a new variable D_i'' is defined for the time spent in physical age interval i.

$$D_i'' = \begin{cases} 0 & (T-f) < A_{i-1} + C \\ (T-f) - (A_{i-1} + C) & \text{if } \ C + A_{i-1} \leqslant (T-f) \leqslant A_i + C\ ; \\ A_i - A_{i-1} & C + A_i \ \ < (T-f) \end{cases} \tag{4:30}$$

$i=1,...,c.$

In principle C is equal to the normal age of entry into the labour market, but in practice the normal (median) age of graduation is used as a proxy variable. A_0 is usually equal to zero. With these specifications employees who enter the labour market at an age below the normal one are not included in the scheme. In order that this group should also be covered, we have to consider a physical age interval to the left of $A_0 + C = C$. To this end a variable D''_0 is defined as*

$$D''_0 = \begin{cases} (T-f) - (A_0 + C) & \text{if } (T-f) < A_0 + C \\ 0 & \text{if } (T-f) \geqslant A_0 + C \end{cases} ; \qquad (4{:}31)$$

The notation for the active age effect in interval i is now changed to γ'_i and the physical age effect is denoted by γ''_i. The new model is now

$$L_{Tfb\cdot} = \exp[\alpha + \beta b + \sum_{i=1}^{c} \gamma'_i D'_i + \sum_{i=0}^{c} \gamma''_i D''_i + \epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tbf}]; \qquad (4{:}32)$$

and

$$E(\epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tfb}) = 0; \qquad (4{:}33)$$

The introduction of the new variable D''_i implies that there is no longer one initial salary. The average salary of those who enter the labour market at an age younger than normal is reduced by γ''_0 for each year below the normal age, as D''_0 is negative. They thus obtain an initial salary smaller than $L_{b(b-C)b\cdot} = \exp[\alpha + \beta b]$ which is the initial salary for those who start work at the normal age. The initial salary for those who are older than normal is assumed to be higher than $L_{b(b-C)b\cdot}$. These relations are exemplified in table 4:3. It is built up under the assumption that the maximum divergence from the normal age of entry into the labour market is one year in either direction. The age intervals are made equal to one year. The salary path for those who start work at a normal age is found in the middle diagonal, for those who start one year late in the lower left diagonal and for those who start one year earlier than normal in the upper right diagonal.
From (4:26) and the following relation

$$T = f + C + \sum_{i=0}^{c} D''_i; \qquad (4{:}34)$$

it follows that

---

* The author is indebted to Harry Lütjohann who suggested the introduction of a D''_0 variable.

Table 4:3. *Expected logarithmic salary path of a labour market cohort (b) with three birth cohorts*

Physical age	Time b	b + 1	b + 2 ...
c − 1	$\alpha+\beta b-\gamma_0''$		
c	$\alpha+\beta b$	$\alpha+\beta b+\gamma_1'$	
c + 1	$\alpha+\beta b+\gamma_1''$	$\alpha+\beta b+\gamma_1'+\gamma_1''$	$\alpha+\beta b+\gamma_1'+\gamma_1''+\gamma_2'$
c + 2		$\alpha+\beta b+\gamma_1'+\gamma_1''+\gamma_2''$	$\alpha+\beta b+\gamma_1'+\gamma_1''+\gamma_2'+\gamma_2''$
c + 3			$\alpha+\beta b+\gamma_1'+\gamma_1''+\gamma_2'+$ $+\gamma_2''+\gamma_3''$
.			
.			
.			

Table 4:4. *Expected logarithmic salary path of a birth cohort with three labour market cohorts*

Active age	Time f + c − 1	f + c	f + c + 1 ...
0	$\alpha'+\beta f-(\gamma_0''+\beta)$	$\alpha'+\beta f$	$\alpha'+\beta f+(\gamma_1''+\beta)$
1		$\alpha'+\beta f+(\gamma_1'-\beta)$	$\alpha'+\beta f+(\gamma_1'-\beta)+(\gamma_1''+\beta)$
2			$\alpha'+\beta f+(\gamma_1'-\beta)+(\gamma_1''+\beta)+$ $+(\gamma_2'-\beta)$
.			
.			
.			

$$b = f + C - \sum_{i=1}^{c} D_i' + \sum_{i=0}^{c} D_i''; \tag{4:35}$$

It is now easily seen that the model (4:32) can be given an alternative formulation

$$L_{Tfb.} = \exp[\alpha' + \beta f + \sum_{i=1}^{c} (\gamma_i' - \beta) D_i' + \sum_{i=0}^{c} (\gamma_i'' + \beta) D_i'' + \epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tfb}]; \tag{4:36}$$

where

$$\alpha' = \alpha + \beta C; \tag{4:37}$$

For a given b (4:32) is the expression for the salary paths of a labour market cohort.★ As those who join the labour market in year b do not all have to be of the same age, there may be more than one path. Analogously, there may be more than one path for each birth cohort. For a given f, (4:36) is the expression for the salary paths of a birth cohort. The first kinds of path are illustrated in table 4:3 and the second in table 4:4 and figure 4:2. There is just one cross section profile which corresponds to these two cohort profiles. It is obtained if b in (4:32) is replaced by (4:26) or if (4:34) is solved for f and substituted into (4:36). With the stochastic elements omitted the result, in logarithmic form, is

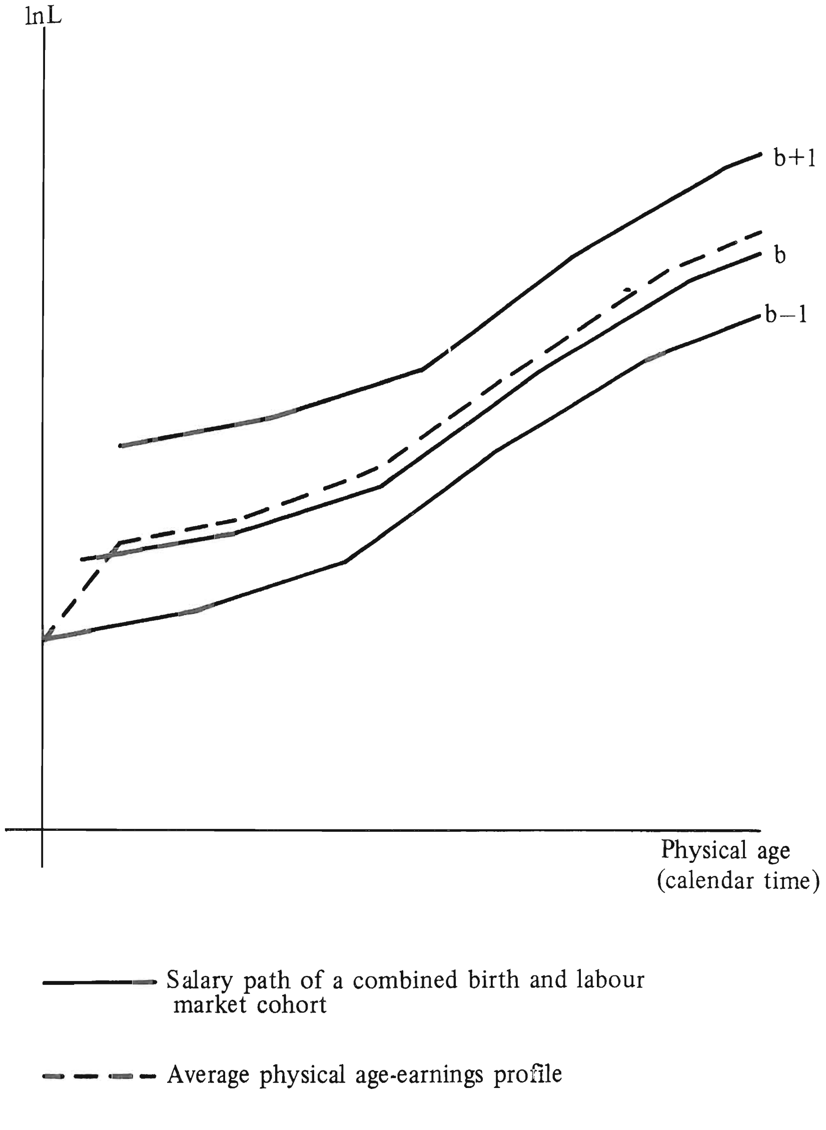
$$E(\ln L_{T_o fb.}) = \alpha + \beta T_o + \sum_{i=1}^{c} (\gamma_i' - \beta) D_i' + \sum_{i=0}^{c} \gamma_i'' D_i''; \tag{4:38}$$

From this expression it is clear that one cross section is enough to estimate the physical age effects. If the average increase in initial salary β is known a priori it is also possible to estimate the active age effects, but if β is not known one needs at least two cross sections.

The three expressions (4:32), (4:36) and (4:38) reveal a certain assymetry between the increments γ_i' and γ_i'', which it is important to understand for a correct interpretation. Suppose we wish to follow the salary paths of a certain labour market cohort. The average salary increase for those who are in the active age interval i and the physical age interval j is $\gamma_i' + \gamma_j''$ from(4:32). Every birth cohort belonging to this labour market cohort obtains the same salary increase due to active age. γ_i' is thus the common salary increase for this labour market cohort. The increase due to physical age is not the same but depends on the birth cohort. γ_j'' is thus a differentiated increase in addition to the common increase.

★ Labour market cohort = employees who enter the labour market in the same year.

Figure 4:2. *Salary paths of one birth cohort with three labour market cohorts*



Suppose now that we prefer instead to follow the salary paths of a given birth cohort. The total salary increase for the same combination of birth and labour market cohorts as above remains of course unchanged. From (4:36) we obtain that $(\gamma_i' - \beta) + (\gamma_j'' + \beta) = \gamma_i' + \gamma_j''$. However, the common increase for all labour market cohorts of this birth cohort is now $(\gamma_j'' + \beta)$, while the additional differentiated increase, differentiated by labour market cohort, is $(\gamma_i' - \beta)$. We thus find that the common increase includes the average increase β, while the differentiated increase does not. To understand this we observe from the cross section (4:38) that β can be interpreted not only as the average increase in initial salaries, but also as the average yearly shift in the cross section profile. β can thus be interpreted as an increase in the general salary level neutral to active and physical age. As a matter of fact, this type of general increase is usually referred to in the literature as economic growth (see for instance Fase [1969]). If salaries are measured in current prices β also includes a general price increase (see section 4.3.6). We may thus say that the model differentiates between three different components of salary increases. With reference to the human capital literature they may be interpreted as a general increase mainly due to economic growth and inflation, an increase due to experience and an increase due to labour productivity associated with physical age. Other interpretations are also possible. β can for instance be seen as the sum of the average negotiated increase and the average salary drift, and $(\gamma_i' - \beta)$ as increases due to promotion.

4.3.3 *A special case*

To use active age as an independent variable is troublesome from the empirical point of view. Salary statistics do not usually contain information about active age. Is it then possible to replace active age by a proxy variable? For some academic groups it is possible to obtain data on when the members of the group received their certificates and degrees. In Sweden most graduate engineers, for instance, go into employment immediately after their studies are finished. Military service is usually done before or during the studies. In this case it seems reasonable to use the time which elapsed since examination as a proxy for active age. However, in most data sources not even this information is available, usually it is only possible to obtain data on birth date or age. What conditions have to be fulfilled to justify the age variable only? The age at which a person starts work is of course first of all determined by his birth date, but education, military service, illness and other personal matters also influence the time of first employment. This influence need not be the same every year. For instance, more and more students stay longer and longer in schools, colleges and universities. In spite of these difficulties it should be possible to handle the problem by considering a group of employees with a conveniently narrow definition. As education seems to be the factor which produces most differentiation, it is natural to consider individuals with the same education.

Suppose now there is such a homogeneous group of employees that every member of the group started work at the same age, i.e.

$$b - f = C; \tag{4:39}$$

From the definitions of the variables D_i' and D_i'' it then follows that

$$\begin{cases} D_0'' = 0; \\ D_i = D_i' = D_i''; \qquad i = 1,\ldots,c \end{cases} \tag{4:40}$$

for every observation. To refer to the example in table 4:3 or table 4:4, these assumptions imply that only the main diagonal in each table exists. Because of the exact multicollinearity between D_i' and D_i'' implied by these assumptions it is not possible to identify both the active and the physical age effects, but only their sum. (4:40) applied to (4:32) gives one formulation of the model in this special case.

$$L_{T(b-C)b\cdot} = \exp[\alpha+\beta b+\sum_{i=1}^{c}(\gamma_i'+\gamma_i'')D_i'+\epsilon_{(b-C)b}+\sum_{t=b+1}^{T}\vartheta_{t(b-C)b}]; \tag{4:41}$$

This expression is nothing but the prototype model (4:23) with a slightly new notation. In the special case it is true that $\gamma_i = \gamma_i' + \gamma_i''$. (4:40) applied to (4:36) gives another formulation,

$$L_{Tf(f+C)} = \exp[\alpha'+\beta f+\sum_{i=1}^{c}(\gamma_i'+\gamma_i'')D_i''+\epsilon_{f(f+C)}+\sum_{t=b+1}^{T}\vartheta_{tf(f+C)}]; \tag{4:42}$$

The first formulation is appropriate when data on active age are available and the second when data on physical age are available. There is of course also a cross section profile which is obtained from (4:38).

$$E(\ln L_{T_0(b-C)b\cdot}) = E(\ln L_{T_0f(f+C)\cdot}) = \alpha+\beta T_0+\sum_{t=1}^{c}(\gamma_i'+\gamma_i''-\beta)D_i; \tag{4:43}$$

In this special case none of the parameters β, γ_i' or γ_i'' are identifiable in one cross section.

If the model formulated as (4:32) or (4:36) is true, then it is possible in the special case considered above to represent an active age-earnings profile by (4:41) and a physical age-earnings profile by (4:42), but in general it is not possible. An investigation of the properties of the estimates obtained when the model (4:42) is applied to data which do *not* fulfil the condition (4:39) is reserved for section 4.3.5. To make this investigation it is necessary first to specify the stochastic properties of the model.

4.3.4 *Stochastic properties of the model*

As the model is applied to average salaries our primary interest is the stochastic properties of the deviation between the average and the expected salary as obtained from (4:32),

$$\ln L_{Tfb\cdot} - E(\ln L_{Tfb\cdot}) = \epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tfb}; \tag{4:44}$$

but in order to obtain realistic assumptions we will investigate how these average deviations are built up of individual deviations from the expected salary.

Individual salaries in a cohort are now treated as realisations of a random process. The expected value of logarithmic individual salaries is given by the non-stochastic parts of the index in (4:32) or (4:36). The deviation of an individual logarithmic *initial* salary from its expectation is denoted by ϵ_{fbi}. As before f stands for birth year, b for the year when the i-th individual joined the labour market and i for the i-th individual. It is assumed that

$$E(\epsilon_{fbi} \mid f,b) = 0; \qquad \text{for all f and b.} \tag{4:45}$$

The mean deviation of n_{bfb} observations is denoted

$$\bar{\epsilon}_{fbi} = \epsilon_{fb}; \tag{4:46}$$

From (4:45) it follows that

$$E(\epsilon_{fb}) = 0; \tag{4:47}$$

The assumptions about second order moments are

$$E(\epsilon_{fbi}\epsilon_{FBI}) = \begin{cases} \sigma^2 & \text{if f=F and b=B and i=I} \\ 0 & \text{otherwise} \end{cases}; \tag{4:48}$$

The variance of ϵ_{fbi} is thus assumed to be constant and independent of when a job is first taken. Alternatively one might assume that the variance increases with time, but there is no strong support for this hypothesis. The assumptions of no correlation between individuals, no correlation between birth cohorts and no autocorrelation are made for convenience. From these assumptions it follows that

$$\mathrm{Var}(\epsilon_{fb}) = \frac{\sigma^2}{n_{bfb}} \; ; \qquad (4{:}49)$$

As the number of young people with the selected educational qualifications entering the labour market has increased during the sample period (table 4:5) the assumption of a constant variance of individual initial salaries implies that the variance (4:49) has decreased. The assumptions made also imply that there is a negative correlation between consecutive average *increases* in initial salary. If the following definition is introduced

$$\epsilon_b = \frac{\sum_f n_{bfb}\epsilon_{fb}}{\sum_f n_{bfb}} \; ; \qquad (4{:}50)$$

it is easy to show that

$$E(\epsilon_{b+1} - \epsilon_b)(\epsilon_b - \epsilon_{b-1}) = -\frac{\sigma^2}{\sum_f n_{bfb}} \; ; \qquad (4{:}51)$$

This is probably a plausible property.

We also consider individual salary increases as random variables with the same expected values as those given by the step function introduced as an approximation to the function G in (4:28). The increase from year t–1 to t

Table 4:5. *Number of young employees in 1954, 1956 and 1969*

Educational qualifications	Age interval –21			–25		
	1954	1956	1969	1954	1956	1969
University degrees						
Engineering				33	78	155
Business & economics				13	11	79
Science				–	6	12
High school certificates						
Engineering I	14	60	205			
Engineering II	20	63	76			
Commerce	–	31	167			

received by the i-th individual, who was born in year f and joined the labour market in year b, deviates from its expected value by ϑ_{tfbi}.★ The assumptions made about this random variable are

$$E(\vartheta_{tfbi} | t,f,b) = 0; \quad \text{for all t, f and b.} \tag{4:52}$$

$$E(\vartheta_{tfbi}\vartheta_{TFBI} | f,b) = \begin{cases} \sigma^2(t-b)(t-f) & \text{if } t=T \text{ and } i=I \\ 0 & \text{if } t \neq T \text{ or } i \neq I \end{cases} ; \tag{4:53}$$

$$E(\vartheta_{tfbi}\vartheta_{TFBI}) = 0; \quad \text{if } b \neq B \text{ and/or } f \neq F$$

(4:52) gives

$$E(\vartheta_{tfb}) = 0; \tag{4:54}$$

and

$$E(\sum_{t=b+1}^{T} \vartheta_{tfb}) = 0; \tag{4:55}$$

What has already been stated as an assumption in (4:33) namely that the expected value of (4:44) is zero follows now from (4:47) and (4:45).

According to (4:53) the variance of the salary increases for a given birth and labour market cohort depends on active and physical age but not on the cohort, i.e. the variance of a certain combination of active and physical age is the same independently of calender year. Fase [1969] makes an even stronger assumption, namely that the variance is constant. This would certainly simplify the model. To get some empirical evidence in favour or against this hypothesis the individual variability will be investigated below as much as the data available permit. Until this investigation eventually justifies a simpler assumption we will keep the more general expression (4:53). The assumptions also prescribe that there is no auto-correlation, no correlation between individuals and no correlation between cohorts. This may be too restrictive but it is introduced in order to make the model simple. Plausible alternatives would be to assume a positive correlation between individuals at the same time and a negative autocorrelation for a given cohort.

Because of (4:53) it is true that

$$\text{Var}(\vartheta_{tfb}) = \frac{\sigma^2_{(t-b)(t-f)}}{n_{tfb}} ; \tag{4:56}$$

★ To refer to the formal representation in section 4.2 it may be observed that the following two identities hold. With a small change in notation to meet the physical age dimension it is true that

$$\epsilon_{fbi} = \epsilon_{fb} + e_{fbi}; \text{ and } \vartheta_{tfbi} = \vartheta_{tfb} + u_{tfbi};$$

and

$$\mathrm{Var}(\sum_{t=b+1}^{T} \vartheta_{tfb}) = \frac{\sum_{t=b+1}^{T} \sigma^2_{(t-b)(t-f)}}{n_{tfb}} \quad ; \tag{4:57}$$

It also follows from the assumptions that all covariances are zero.

(4:57) shows that the variance of the average salary for a combined birth and labour market cohort increases by (active) age. The same thing of course also holds for individual salaries. This is a property which closely agrees with previous experience from earnings data. Empirically we know that the higher the average salary and the older the employees, the greater the dispersion (Hill [1959], Morgan [1962]). One important explanation for this is that most employees start at about the same job level at approximately the same salary. Some then obtain promotion to more responsible and higher paid jobs, while others do not, or at least not as quickly. This suggests that the variance of salary increases is highest in the age interval where promotion differentiates the salary increases most. During the first five to ten years in the labour market most employees with academic or high school education obtain a more or less normal promotion to middle level jobs. It is, however, more difficult to obtain further promotion to the relatively few jobs available at the top levels. It is therefore a plausible hypothesis that the variance is for this reason the highest somewhere in the age interval 30–45 years.

The published salary statistics from the Swedish Employers' Confederation admit a rough check of this hypothesis. Table 4:6 contains semi-interquartile ranges calculated from the published tables of graduate engineers and high school engineers. They exhibit a variability which clearly increases by age. However, a warning must be given against this interpretation. The age intervals used differ in length and since the average salary increases by age, the dispersion may be expected to be higher in wide intervals than in narrow ones. As wide intervals are used at the end of the age distribution, table 4:6 may exaggerate the increase in dispersion. Furthermore, the changes in average salary from one year to another contribute to the measured variability for an age interval independently of its length. The semi-interquartile ranges contain a component which is the variability between averages. They thus tend to overestimate the individual dispersion. The measures of dispersion in table 4:6 need to be standardized for these effects before a proper comparison can be made. This is done below.

The SAF data are only grouped by physical age. To simplify the calculations the special case model (4:42) will be used although the data do not satisfy the condition of the special case. For our purpose this is probably no serious disadvantage. The following notation is introduced

$$\mu_{Tf} = E(\ln L_{Tf(f+C)\cdot}); \tag{4:58}$$

$$\gamma_i = \gamma_i' + \gamma_i'' \tag{4:59}$$

$\sigma^2_{(T-f)}$ is the variance of individual salaries at the age (T−f).

S_i^2 is the theoretical variance of all individual salaries in the age interval $((A_{i-1} + C) - (A_i + C)$ of a cross section.

Assume that the age distribution *inside* each age interval is uniform and denote the number of employees at *each* age in the i-th interval n_i. It is then possible to write S_i^2 as follows

$$S_i^2 = \frac{1}{n_i} \sum_{(T-f)} \sigma^2_{(T-f)} + \sum_f \sum_k \left(\frac{\mu_{Tf} - \mu_{Tk}}{n_i}\right)^2 ; \tag{4:60}$$

The first term of this expression is the mean of the individual variances in age interval i, and the second term is the sum of all $\binom{n}{2}$ possible squared differences $(\mu_{Tf} - \mu_{Tk}) \frac{1}{n_i}$. From (4:42) and (4:59) it follows that

$$\mu_{Tf} - \mu_{Tk} = (\gamma_i - \beta)(f - k); \tag{4:61}$$

in the i-th interval. If S_i^2, γ_i and β were known the mean variance of individual salaries could be calculated. They are not, but it is possible to obtain crude estimates of S_i^2 from table 4:6 and preliminary estimates of γ_i and β, based on simple assumptions about the stochastic components.★ The results from calculations with these estimates are presented in table 4:7. The variance component due to unequal age intervals and salary increases is small compared with the total variance. The remaining individual variability increases by age.

The estimates in table 4:7 are interesting not only because they give some insight into the individual variability by age, but also because they indicate the magnitude of the individual variability. As natural logarithms have been used, the standard deviations in the table multiplied by 100, can approximately be interpreted as percentages. We may for instance say that for employees 28–29 years old the individual dispersion is 14.7 % of the average salary in this age interval. The magnitudes may be more sensitive to the approximations made than the association with age, but if the present estimates are trusted, they seem to indicate that the individual dispersion for young employees is of the same magnitude as their average salary increases (see table 4:11), while it is higher for older employees.

★ The model is estimated by a procedure in three steps. In the first step the variance of the random residual (4:44) is treated as a constant and γ_i and β are estimated by the method of ordinary least squares. These estimates are in the second step used to specify a new moment matrix (section 4.3.4). Finally this new matrix is used for a reestimation of the parameters of the model, (section 4.3.6).

Table 4:6. *Semi-interquartile ranges of logarithmic salaries*

Education/cross section per year	Physical age											
	20–21	22–23	24–25	26–27	28–29	30–31	32–34	35–39	40–44	45–49	50–59	60–
Graduate engineers												
1956			0.039	0.071	0.106	0.113	0.120	0.138	0.166	0.210	0.216	0.292
1962			0.039	0.071	0.092	0.104	0.115	0.117	0.140	0.150	0.180	0.237
1969			0.051	0.062	0.104	0.104	0.113	0.117	0.127	0.134	0.164	0.196
High school engineers I												
1956	0.037	0.067	0.067	0.104	0.111	0.111	0.127	0.150	0.182	0.184	0.203	0.235
1962	0.044	0.060	0.067	0.083	0.092	0.104	0.117	0.129	0.152	0.173	0.187	0.242
1969	0.058	0.067	0.092	0.101	0.104	0.115	0.117	0.140	0.140	0.152	0.182	0.212

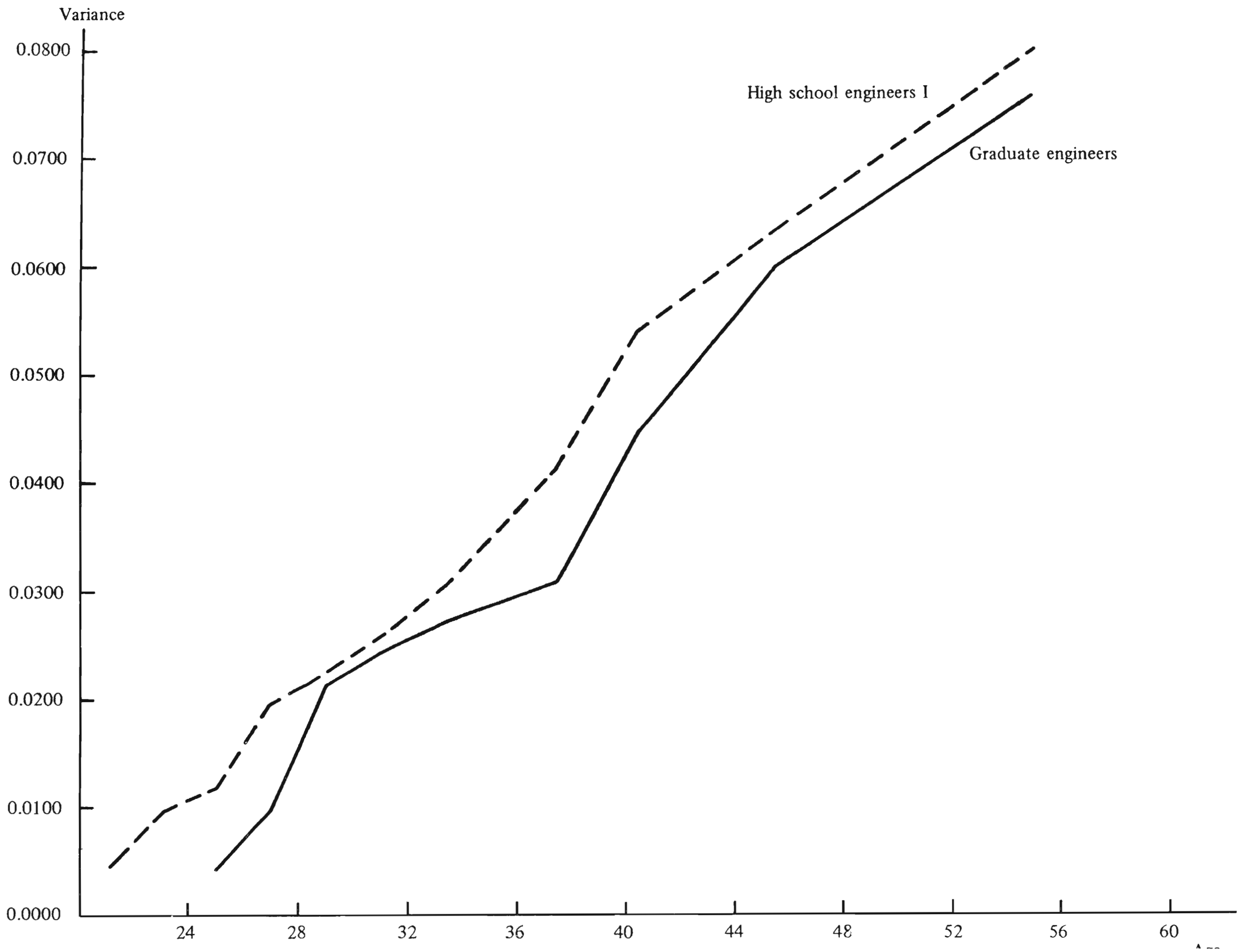
Note: The semi-interquartile range is obtained as 1/2 ($\ln Q_3 - \ln Q_1$), where Q_1 and Q_3 are the first and the third quartile salary. As natural logarithms have been used, the semi-interquartile ranges multiplied by 100 can be interpreted as percentages.

Table 4:7. *Individual salary variability*

	Physical age										
	20–21	22–23	24–25	26–27	28–29	30–31	32–34	35–39	40–44	45–49	50–59
Graduate engineers											
Variance due to unequal age intervals and salary increases			0.0002	0.0006	0.0008	0.0009	0.0020	0.0029	0.0008	0.0002	0.0007
Variance due to individual variability			0.0039	0.0096	0.0217	0.0244	0.0276	0.0310	0.0446	0.0598	0.0760
Standard deviation, ditto			0.062	0.098	0.147	0.156	0.166	0.176	0.211	0.245	0.276
High school engineers I											
Variance due to unequal age intervals and salary increases	0.0002	0.0003	0.0005	0.0006	0.0006	0.0005	0.0009	0.0015	0.0006	0.0000	0.0002
Variance due to individual variability	0.0044	0.0089	0.0118	0.0196	0.0222	0.0261	0.0308	0.0418	0.0542	0.0635	0.0799
Standard deviation, ditto	0.066	0.094	0.109	0.140	0.149	0.162	0.175	0.204	0.233	0.252	0.283

Note: The results in this table are obtained by an approximation. S_i in expression (4:60) is estimated by an average of three semi-interquartile ranges from table 4:6 divided by the factor 0.6745 which is obtained from a table of the normal distribution. As natural logarithms have been used,the standard deviations multiplied by 100 can be interpreted as percentages.

Figure 4:3. *Individual salary variability by physical age*



In figure 4:3 the variances due to individual variability are plotted against age. There is a slight indication of a smaller increase in variability between 30 and 40 years for graduate engineers and a few years earlier for high school engineers, and a higher increase than average immediately before and after this age interval. This may indicate a somewhat smaller variance of salary increase between 30 and 40 years (25 and 35 for high school educated employees) than among younger and older employees. This is a result almost opposite to that expected a priori. The strategy to adopt is therefore to disregard possible variations in the variance of individual salary *increases* and to base the specification instead on the fact that the general shape of the curves can very well be approximated by two linear functions, one for graduate engineers, which intersects the age axis at 24–25 years and one for high school engineers which intersects at about 20 years. The slopes of the linear functions can be interpreted as estimates of constant variances of salary increases. The approach chosen by Fase [1969] has thus got some justification.

The variance of individual salary increases (4:53) is now assumed to be constant and independent of both active and physical age. It is also assumed that this variance equals the variance of initial salaries (4:48) and that the age dependent salary increases are uncorrelated with the initial salaries. The last assumption may seem too restrictive, but as the model is applied to relatively homogeneous educational groups it is acceptable. From these assumptions and from (4:57) it follows that the variance of the average deviation (4:44) is

$$D^2 = \text{Var}(\epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tfb}) = \frac{(T-b+1)\sigma^2}{n_{tfb}}; \tag{4:62}$$

To investigate the sensitivity of the estimates to alternative specifications, the following alternatives were also tried

$$D^2 = \frac{(T-f-C+1)\sigma_1^2}{n_{tfb}}; \tag{4:63}$$

$$D^2 = \frac{\sigma_2^2}{n_{tfb}}; \tag{4:64}$$

$$D^2 = \sigma_3^2; \tag{4:65}$$

In the first alternative the variance increases by physical age rather than by active age, in the second it only depends on the number of employees which implies a certain negative correlation between successive salary increases. This is also true for the third alternative in which the variance is in addition independent of the number of employees.

4.3.5 *Specification errors*

We now return to the problem in which a physical age-earnings profile is estimated from data which lack information about active age and do not satisfy (4:39), i.e. all employees do not take their first job at the same age. At the end of section 4.3.3 it was stated that a physical age-earnings profile could not in general be represented by model (4:42). Suppose now that this model is used as though everybody started work at the same age although it is not true. What properties do our estimates of (4:42) have in this case?

To give a first answer to this question we observe that (4:42) is now applied to average salaries, $L_{Tf..}$, each of which is calculated for individuals of the same physical age but not necessarily of the same active age.

$$\ln L_{Tf..} = \frac{\sum_b n_{Tfb} \ln L_{Tfb.}}{\sum_b n_{Tfb}} ; \qquad (4:66)$$

As before n_{Tfb} is the number of employees in the combined birth and labour market cohort in year T. If (4:36) is the »true» model, the »true» expression for the physical age-earnings profile can be derived if (4:36) is substituted into (4:66). It is thus a weighted average of salary paths for combined birth and labour market cohorts, (see the dashed curve in figure 4:2).

The general expression for this physical age-earnings profile is awkward and does not add much new insight. We use instead the example in table 4:4 which is a representation of the salary paths of a birth cohort with three labour market cohorts. Suppose there are n_1 employees in the labour market cohort (b−1) = = (f+C−1), n_2 in the cohort b = (f+C) and n_3 in (b+1) = (f+C+1). The physical age-earnings profile (in logarithms) is obtained if for each physical age (each column) we calculate a weighted average from the salaries of each labour market cohort (each row). The weights are proportional to the number of employees in each labour market cohort. The stochastic components are not written out in table 4:4 but they can easily be obtained from (4:36). For instance, for the first age when all labour market cohorts are represented, C+1, the mean logarithmic salary is

$$\ln L_{(f+C+1)f..} = \alpha' + \beta f + (\gamma_1' - \beta) + (\gamma_1'' + \beta) + \frac{n_1\gamma_2' - n_3\gamma_1'}{n_1 + n_2 + n_3} +$$
$$+ \frac{n_1[\epsilon_{f(f+C-1)} + \vartheta_{(f+C)f(f+C-1)} + \vartheta_{(f+C+1)f(f+C-1)}]}{n_1 + n_2 + n_3} +$$
$$+ \frac{n_2[\epsilon_{f(f+C)} + \vartheta_{(f+C+1)\,f(f+C)}] + n_3\epsilon_{f(f+C+1)}}{n_1 + n_2 + n_3} ; \qquad (4:67)$$

From our previous assumptions about the random variables in (4:36) it follows that the composite stochastic variable in (4:67) has zero expectation and the variance

$$\mathrm{Var}[\ln L_{(f+C+1)f..} - E(\ln L_{(f+C+1)f..})] = \frac{2\sigma^2}{n_1+n_2+n_3} + \frac{(n_1-n_3)\sigma^2}{(n_1+n_2+n_3)^2}\,; \quad (4:68)$$

Model (4:42) gives the following representation of the same average salary $L_{(f+C+1)f..}$

$$\ln L_{(f+C+1)f..} = \alpha' + \beta f + (\gamma_1'-\beta) + (\gamma_1''+\beta) + \epsilon_{f(f+C)} + \vartheta_{(f+C+1)f(f+C)}; \quad (4:69)$$

$$E(\epsilon_{f(f+C)} + \vartheta_{(f+C+1)f(f+C)}) = 0; \quad (4:70)$$

$$\mathrm{Var}(\epsilon_{f(f+C)} + \vartheta_{(f+C+1)f(f+C)}) = \frac{2\sigma^2}{n_1+n_2+n_3}\;; \quad (4:71)$$

A comparison between (4:67), (4:68) and (4:69) – (4:71) shows two specification errors involved in the application of (4:42). Firstly a »variable» is omitted, an element of which is the last non-stochastic term in (4:67), and secondly the residual variance is wrongly specified. If the estimation method is the method of ordinary least squares the last error does not introduce a bias, but the first error in general does. The magnitude of the bias depends on the correlation between the omitted variable and the independent variables in (4:42). If it is uncorrelated with all of them there is no bias. This is, however, rather unlikely. If the assumptions behind the curve TT in figure 4:1 are true, the difference between consecutive salary increments should be positive for young and negative for old employees. A negative correlation with D_i'' is thus more likely than no correlation.

In practice model (4:42) is applied with a small modification. In its original form it does not give any representation for average salaries of employees who are younger than the normal age on joining the labour market. In order to adapt the model so as to include these observations also the variable D_0'' is added as an independent variable. With an easily understood modification of the notation this new model variant is written as

$$\ln L_{Tf..} = \tilde{\alpha} + \tilde{\beta}f + \sum_{i=0}^{c} \tilde{\gamma}_i D_i'' + \tilde{\vartheta}_{Tf}\;; \quad (4:72)$$

The error component $\widetilde{\vartheta}_{Tf}$ is treated as a stochastic variable with zero expectation and for different values of T and f the components are treated as if they were uncorrelated. A natural variance specification in the age intervals i=1,...,c is a specification analogous to (4:63). The specification for the interval i=0 is less obvious. In practice (4:64) is used with $\sigma_1^2 = \sigma_2^2$.

To make a comprehensive analysis of the specification errors it would be necessary to find the expected values of the least squares estimates of the parameters in (4:72). This requires knowledge about the number of employees in each labour market cohort n_{Tfb}, and it would lead to tedious derivations and awkward expressions. We prefer instead to investigate the bias under two simplifying assumptions. Contrary to what is possible in the practical situation described in the beginning of this section, it is assumed that (4:72) is applied to average salaries for each combination of active and physical age, $L_{Tfb.}$, instead of to averages for all employees of the same age, $L_{Tf..}$. In addition it is assumed that the method of ordinary least squares is used in spite of the variance specification suggested.

To simplify the exposition a new compact notation is used. Three new matrices are defined, X_1, X_2 and X_3. The unit regressor, the f-regressor and the D_0''-regressor are combined in the matrix X_1, the D_i'-regressors in X_2 and the D_i''-regressors, with the exception of D_0'', in X_3. The vector of logarithmic average salaries is denoted ℓ, the vector of stochastic components in the »true» model (4:36) by ν and in the model actually applied, (4:72) by ξ. We also need five new parameter vectors $g_1' = \{\alpha', \beta, \gamma_0'' + \beta\}$, $g_2' = \{\gamma_1' - \beta, ..., \gamma_c' - \beta\}$, $g_3' = \{\gamma_1'' + \beta, ..., \gamma_c'' + \beta\}$, $h_1 = \{\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\gamma}_0\}$ and $h_3 = \{\widetilde{\gamma}_1, ..., \widetilde{\gamma}_c\}$. The »true» model is now

$$\ell = \{X_1 \mid X_2 \mid X_3\} \begin{Bmatrix} g_1 \\ \hline g_2 \\ \hline g_3 \end{Bmatrix} + \nu; \tag{4:73}$$

and the applied model is

$$\ell = \{X_1 \mid X_3\} \begin{Bmatrix} h_1 \\ \hline h_3 \end{Bmatrix} + \xi; \tag{4:74}$$

It can be shown by standard regression theory that the expectation of the ordinary least squares estimates of h_1' h_3' in (4:74) is

$$E \begin{Bmatrix} \hat{h}_1 \\ \hline \hat{h}_3 \end{Bmatrix} = \begin{Bmatrix} g_1 \\ \hline g_3 \end{Bmatrix} + B_{2.13} g_2; \tag{4:75}$$

where $B_{2.13}$ is a matrix of regression coefficients from the regression of X_2 on X_1 and X_3. Provided the inverses of $(X_1'X_1)$ and $(X_3'X_3)$ exist this expression can be rewritten as follows

$$E\begin{Bmatrix}\hat{h}_1\\ \hat{h}_3\end{Bmatrix} = \begin{Bmatrix} I & 0 & 0\\ 0 & I & I\end{Bmatrix}\begin{Bmatrix}g_1\\ g_2\\ g_3\end{Bmatrix} + \begin{Bmatrix}(X_1'X_1 - X_1'X_3(X_3'X_3)^{-1}X_3'X_1)^{-1}(X_1' - X_1'X_3(X_3'X_3)^{-1}X_3')\\ (X_3'X_3 - X_3'X_1(X_1'X_1)^{-1}X_1'X_3)^{-1}(X_3' - X_3'X_1(X_1'X_1)^{-1}X_1')\end{Bmatrix}(X_2 - X_3)g_2; \tag{4:76}$$

In the special case discussed above when all employees take their first job at the same age, it is true that $X_2 - X_3 = 0$, but it is also true that the regressor D_0'' is a zero vector. Thus, the inverse of $(X_1'X_1)$ does not exist and the expression (4:76) is not valid. Suppose instead for a moment that D_0'' is omitted from both models. The inverse of the new matrix $(X_1'X_1)$ exists. We then find that the last term of (4:76) is zero and that we obtain unbiased estimates of the parameters in (4:42) if $X_2 = X_3$, i.e. if the data satisfy the condition for the special case we obtain unbiased estimates of (4:42).

Suppose now that the special case is not present and that D_0'' is not the zero vector. From (4:76) it follows that the more X_2 deviates from X_3, the larger the bias, as defined by the last term in the expression. It is difficult to give any general rules for the sign and magnitude of the bias without knowledge of the X matrices and g_2. To give some indications a simple numerical example is worked out. It contains two cross sections T_1 and T_2, two active age intervals and three physical age intervals, each of one year. In the same way as in the previous example there are three birth cohorts for each labour market cohort. The regressors are as follows

	(α')	f	D_1'	D_2'	D_0''	D_1''	D_2''
T_1	1	4	0	0	−1	0	0
	1	3	0	0	0	0	0
	1	2	0	0	0	1	0
	1	3	1	0	0	0	0
	1	2	1	0	0	1	0
	1	1	1	0	0	1	1
	1	2	1	1	0	1	0
	1	1	1	1	0	1	1
T_2	1	5	0	0	−1	0	0
	1	4	0	0	0	0	0
	1	3	0	0	0	1	0
	1	4	1	0	0	0	0
	1	3	1	0	0	1	0
	1	2	1	0	0	1	1
	1	3	1	1	0	1	0
	1	2	1	1	0	1	1

(4:77)

With these regressors the expression (4:76) becomes

$$E\left\{\begin{array}{c}\hat{h}_1\\ \hdashline \hat{h}_3\end{array}\right\} = \left\{\begin{array}{c}\alpha'\\ \beta\\ \gamma_0'' + \beta\\ \hdashline \gamma_1' + \gamma_1''\\ \gamma_2' + \gamma_2''\end{array}\right\} + \left\{\begin{array}{rr}0.872 & 5.053\\ -0.011 & -0.531\\ 0.373 & 0.415\\ \hdashline -0.837 & -0.057\\ -0.021 & -0.348\end{array}\right\} \left\{\begin{array}{c}\gamma_1' - \beta\\ \hdashline \gamma_2' - \beta\end{array}\right\}; \tag{4:78}$$

If for instance

$$\left\{\begin{array}{c}\gamma_1' - \beta\\ \gamma_2' - \beta\end{array}\right\} = \left\{\begin{array}{r}0.06\\ -0.02\end{array}\right\} \tag{4:79}$$

the bias vector is

$$\{-0.049, 0.010, 0.014, -0.049, 0.006\} \tag{4:80}$$

This example gives an idea of the magnitude of the bias. In practice the observations probably conform to the special case better than in the example, i.e. the age distribution is more concentrated around the normal (median) age of entry into the labour market than in the example. If this is so, the bias vector (4:80) may exaggerate the bias likely to occur in practice. Nevertheless the example indicates that it is not advisable to disregard the bias since the absolute value of the fourth bias element in (4:80) is almost half as large as a realistic value of the corresponding parameter. If it is at all possible to generalize, the example shows that the salary increases of young employees are underestimated rather much, while the bias for older employees is less important.

The same kind of specification error analysis could be accomplished if active age-earnings profiles were estimated by (4:41) on data which do not follow the special case. Such an analysis would give results similar to those already obtained.

Before the discussion of the improper application of the model (4:42) is concluded, yet another source of bias should be mentioned. In all the analysis above, the normal age of entry into the labour market was assumed constant for a given education. If this assumption is not satisfied the result may become a bias. The problem is not important for the model (4:36), at least as long as γ_i'' is independent of the age distribution, but for the application of (4:42) it may be important, even when the condition for the special case is satisfied. However, it is only a potential problem and no real problem for the sample period, because the median examination age has been stable (see table 4:8). For this reason it does not seem necessary to elaborate the problem.

Table 4:8. *Average age when a degree or certificate is obtained in 1960–1970*

Educational qualifications		1961/62	1962/63	1963/64	1964/65	1965/66	1966/67	1969/70
University degrees								
Engineering	$\bar{X}$	26.4	25.9	26.2	26.5	26.0		26.0
(Civilingenjör)	S	2.8	2.5	2.8	2.7	2.7		1.9
Business & economics	$\bar{X}$			26.8				26.3
(Civilekonom)	S			2.4				2.9
Science	$\bar{X}$			26.2				26.1
(Naturvetare, FK, FM)	S			3.8				3.5
High school certificates								
Engineering I	$\bar{X}$						20.7	20.1
(Läroverks-ingenjör)	S						2.4	1.7
Engineering II	$\bar{X}$						23.0	24.2
(Instituts-ingenjör)	S						3.2	3.9
Commerce	$\bar{X}$			20.8			20.1	19.5
(Gymnasie-ekonom)	S			2.5			1.8	2.5

$\bar{X}$ = mean age; S = standard deviation.

Data sources and comments:

University degrees

Engineering: 1961/62–1965/66. Age when the degree was obtained. Unpublished data from »Teknikerundersökningen», Central Bureau of Statistics 1969/70. Age in 1970 of those who obtained a degree in 1969/70. Unpublished, Central Bureau of Statistics.

Business & science: Unpublished, Central Bureau of Statistics 1963/64. Age when the degree was obtained in 1969/70. Age in 1970 of those who obtained a degree in 1969/70.

High school certificates

Age is measured during the autumn of the last year at school. Statistical Reports U 1964:13, 1968:15, 1970:20; Central Bureau of Statistics.

Engineering II: Full-time students at public (kommunala) and private technical schools.

In conclusion the specification error analysis shows that if the primary interest is to estimate the parameters of the model (4:32) or (4:36), it is not a proper procedure to estimate a pure physical age-earnings profile or active age-earnings profile without an evaluation of the bias and with possible adjustments. This does not, however, exclude applications in which the only interest is to make a description of an age-earnings profile in the physical age dimension or in the active age dimension without any reference to the active and physical age effects as they are defined in (4:32) or (4:36). In those cases (4:72) and (4:41) respectively are of course appropriate.

The last paragraphs of this section are devoted to a different kind of specification error. To use least squares estimation it is not necessary to specify a particular distribution for the residuals, but if it were possible to do so with a reasonable degree of approximation, more precise statements could of course be made about the estimates. This is, however, not the main reason for considering a lognormal distribution for the residuals. There is instead the following problem. $L_{Tfb.}$ is defined as a geometrical average, but the SAF tables only give median salaries. Is it then possible to justify an estimation from median salaries?

The lognormal distribution is commonly used as an income distribution and it does not seem unrealistic to assume that the salary distribution conditional on cohort and age (possibly both physical and active age) can be approximated by this distribution (see for instance Hill [1959]).

If the stochastic variable x follows a lognormal distribution, then ln(x) follows a normal distribution with expected value and standard deviation, say μ and σ respectively. From the well known properties of the lognormal distribution it then follows that $\exp[\mu]$ is the median of the lognormal distribution, while the expected value is $\exp[\mu + \frac{\sigma^2}{2}]$.

If L_{Tfbi} follows a lognormal distribution with the parameters μ and σ and we look upon $L_{Tfb.}$ as a sample average, this will also follow a lognormal distribution, but with the expected value $\exp[\mu + \frac{\sigma^2}{2n_{fb}}]$. ★ For large n_{fb} the dif-

---

★ y is independently $N(\mu, \sigma)$

$x = e^y$ is then lognormal with parameters μ and σ, and $E(x) = \exp[\mu + \frac{\sigma^2}{2}]$;

$$E(\prod_{i=1}^{n} x_i^{\frac{1}{n}}) = E(e^{\frac{\Sigma y_i}{n}}) = E(e^{\bar{y}}) ;$$

$\bar{y}$ is $N(\mu, \frac{\sigma}{n})$;

$$\because E(e^{\bar{y}}) = \exp[\mu + \frac{\sigma^2}{2n}] ;$$

ference between the theoretical median salary and this expected value is small. Furthermore, if n_{fb} is sufficiently large, the empirical geometrical average and the empirical median salary will not deviate much from their respective theoretical values $\exp[\mu + \frac{\sigma^2}{2n_{fb}}]$ and $\exp[\mu]$. No great error should thus be committed if the empirical geometric average is replaced by the observed median salary.

4.3.6 *Estimation*

The data used to estimate the models are obtained from SAF's published tables of median salaries and from the Swedish Association of Graduate Engineers. From the last source geometrical averages have been obtained by age, year of graduation and calendar year. The SAF tables only give median salaries by physical age. It is important to note that the data do not cover salaries of identical individuals but are several cross sections combined in one sample. The model does not require salary data for identical individuals, all that is necessary is that the salary paths of all individuals covered by the estimations can be represented by one and the same model.

The model can be estimated from data both in current and constant prices. In principle the estimates obtained in the two cases will naturally differ as regards both value and expectation. Suppose that price development can be represented by an exponential trend

$$P_T = \exp[\pi T + \zeta_T] ; \tag{4:81}$$

P_T is a price index with the base 1 for T=0. The expression (4:81) is to be taken as a smoothing formula. The constant π can for instant be calculated by the method of least squares and interpreted as the average price increase during the sample period. The ζ_T's are observed residuals from the fitted trend. If they are least squares residuals, they are uncorrelated with T. From (4:18) and (4:32) it now follows that

$$L^R_{Tfb.} = \exp[\alpha + (\beta-\pi)b + \sum_{i=1}^{c} (\gamma_i' - \pi)D_i' + \sum_{i=0}^{c} \gamma_i'' D_i'' + $$

$$+ \epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tfb} + \zeta_T]; \tag{4:82}$$

If ζ is also uncorrelated with b, D_i' and D_i'', then a least squares estimation from data in constant prices gives unbiased estimates of the »parameters» in

(4:82) conditional on P_T for all values of T in the sample period. Even if the price residuals are not perfectly uncorrelated with the independent variables, this will be valid as a reasonable good approximation during the sample period. If this is true it is not very important what kind of data are used. It is possible to go from one set of estimates to the other by a simple adjustment by the average price increase. For the sample period the average increase in the consumer price index is about 4 % (see table 4:12).

In an economic model which aims at an explanation of the causes underlying salary increases, an exponential trend would probably be too simple a model of price increases. As will be suggested (but not done) in the next section, more refined explanations of salary changes can be added to the model. It seems natural therefore to explain salary changes in current prices by changes in the consumer price index and in this way to incorporate the price changes into the model instead of using preadjusted data. The approach chosen in this study is to use data in current prices.

(4:82) also clearly reveals the asymmetry between salary increases due to active age and salry increases due to physical age. The γ_i'''s are, so to speak, already real increases by definition, while the γ_i''s are not. This is important for a comparative interpretation of the estimates.

Three variants of the model, namely (4:32), (4:41) and (4:72), have been estimated from CF data for graduate engineers specializing in electrical engineering (SER) or mechanical engineering (SMR). For each variant alternative variance specifications have been tried. Model (4:32) is combined with all four specifications (4:62)–(4:65), model (4:41) with all but (4:63) and model (4:72) with all but (4:62). C in (4:63) is put equal to 24. This value is obtained by inspection of figure 4:3.

When the residual variance is assumed to follow any of the specifications (4:62)–(4:64) the estimation method used is that of weighted least squares, while ordinary least squares is used when the variance is assumed to be constant, (4:65).

The salary data are recorded by age intervals and the salaries of the youngest are grouped in the interval 25 years and below. The estimation is made under the assumption that the average age in this interval can be approximated by 25 years. As C is also put equal to 25 (except in the variance specification) this means that γ_0'' is assumed to be zero and that the variable D_0'' is omitted. The fact that data are grouped in intervals makes it impossible to obtain an observation for each age year. Each interval instead gives one observation and the value of the age variable is put equal to the midpoint of the interval. This is probably no great disadvantage as there are rather many intervals. The number of intervals in the data tables is about twice as large as the number of intervals defining the variables D_i' and D_i'' in the model. There is thus usually more than one observation in each model interval.

Estimates of the three model variants applied to SER data are presented in table 4:9 and the results from SMR data in table 4:10. The multiple correlation coefficients show that the specification with a constant residual variance (4:65) is inferior to the other three specifications. The closest fit is obtained when the variance is proportional to physical age, (4:63), but the difference between this specification and (4:62) and (4:64) is so small that the choice between them can hardly be based on difference in fit. The specification primarily derived from the model made the variance proportional to active age, (4:62), and there is no great reason to reject it. As the results from SMR data closely agree with those from SER data only selected estimates are given in table 4:10.

The annual increase in initial salaries is estimated at 5–6 % irrespective of model, variance specification and data set. After deduction of the general price increase there remains 1–2 % real increase. The total salary increase for a given combination of active and physical age is obtained if the two age effects are added. For instance, from the first row in table 4:10 we find that an engineer specializing in electrical engineering who graduated 5–9 years ago and is 35–39 years old obtains on average 8.7 + 2.0 = 10.7 % salary increase per year. As the estimates are obtained from data in current prices the figure of 8.7 % due to active age includes a general salary increase which can primarily be explained by price increases and economic growth, (see section 4.3.2, p.92). A measure of this general increase is obtained in the estimate of β, i.e. 5.5 % in this case. The difference between the effect due to active age and the general increase, 3.2 %, may be interpreted as an increase due to experience. A comparison with the effect due to physical age, 2 %, shows that experience means more than physical age as regards salary increases. This conclusion is also supported by the better fit of the active age model (4:41) than of the physical age model (4:72). The physical age effect is, however, important enough not to be neglected.

The increases due to active age decrease as the engineer stays in the labour market. This agrees with Becker's human capital theory except for the very beginning of one's career (see figure 4:1). According to Becker so much of the first years after graduation is devoted to on-the-job training that the salary increases very slowly during these years. No tendency of a lower rate of increase during the first years on the labour market can be observed in tables 4:9–4:10. However, this effect may be hidden in the first age class 0–4 years and a year-by-year analysis may reveal a low rate for the very first years.

The salary increases associated with physical age also decrease as the employee grows old. After the age of 40–45 years physical age does not contributes anything to the salary increase, on the contrary there is a tendency towards a decrease. This may perhaps be seen as a premium for higher efficiency in work among young engineers than among middle aged and old ones.

Tables 4:9–4:10 also give a possibility of indirectly checking the assumption of a constant graduation age. The sum of the estimates of the physical age effects and the corresponding active age effects in (4:32) do not of course equal the estimates of (4:41) and (4:72) exactly, but the divergence is not large enough to disturb the general shape of the profiles. This is in particular true for (4:41). In the lower age brackets the sum of the estimates of γ_i' and γ_i'' in model (4:32) is higher than the corresponding estimate in (4:72), in particular when the residual variance is assumed to be constant. This result can be compared with the bias in the example previously worked out. As the reader will recall, the bias was negative in the first age interval.

What generalizations to »similar» data sets can be made from these results is a matter of opinion. They may give some reassurance that bias will not completely invalidate the following application of model (4:72) to SAF data. At any rate, what follows below can be seen as a description of the sample period.

In order to avoid misinterpretations of the estimates presented below, the reader is reminded of the selection bias in the SAF data (see chapter 2). The statistics from SAF do not cover employees at the top management level. Older employees who remain in the statistics are therefore those who have not obtained a promotion. The practice of not submitting salaries of employees at management level may, in small companies, even result in non-response at a relatively low job level. The result of this selection is an underestimation of salary increases and salary levels for middle aged and old employees. There are no statistics collected to exhibit the salaries at management level and it is therefore difficult to know the importance of this selection bias. It should be most severe among graduate engineers and business economists who are most frequent at management level, and less severe among high school educated employees. An attempt has been made in Klevmarken [1968c] to use statistics from CF as a comparison. There are reasons to believe that the selection is not equally strong in this data source, and the estimates obtained of the rates of increase in earnings are also a few percentage units higher in the age bracket 35 years and over than the estimates based on SAF data.★ In an attempt to throw more light on this

★ In this comparison the CF data used cover engineers employed by members *and* non-members of SAF, but even if non-members are eliminated a considerable difference remains as the analysis in chapter 2 shows.

Table 4:9. *Estimated salary profiles of graduate engineers specializing in electrical engineering (SER)*

Model and variance specification	Annual increase in initial salary	Annual increase in active age interval 0–4	5–9	10–14	15–19	20–24	25–29	Annual increase in physical age interval –29	30–34	35–39	40–44	45–49	50–54	55–59	R
(4:32)															
$\frac{(T-b+1)}{n_{Tfb}}\sigma^2$	0.055	0.107	0.087	0.079	0.073	0.073	0.070	0.024	0.025	0.020	0.002	–0.008	–0.037	–0.006	0.9789
	(0.001)	(0.002)	(0.002)	(0.004)	(0.007)	(0.011)	(0.023)	(0.002)	(0.002)	(0.004)	(0.006)	(0.010)	(0.016)	(0.038)	
$\frac{(T-f-23)}{n_{Tfb}}\sigma^2$	0.056	0.114	0.092	0.081	0.075	0.074	0.071	0.022	0.019	0.016	0.001	–0.010	–0.039	–0.003	0.9883
	(0.001)	(0.002)	(0.002)	(0.003)	(0.005)	(0.008)	(0.015)	(0.002)	(0.002)	(0.003)	(0.005)	(0.007)	(0.012)	(0.030)	
$\frac{\sigma^2}{n_{Tfb}}$	0.058	0.113	0.092	0.086	0.074	0.079	0.071	0.026	0.019	0.016	0.001	–0.010	–0.039	–0.002	0.9871
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.006)	(0.009)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.008)	(0.017)	
σ^2	0.059	0.105	0.095	0.076	0.084	0.061	0.110	0.025	0.020	0.024	0.000	–0.009	–0.042	0.003	0.9412
	(0.001)	(0.003)	(0.003)	(0.006)	(0.012)	(0.017)	(0.022)	(0.004)	(0.002)	(0.005)	(0.009)	(0.012)	(0.016)	(0.022)	
(4:41)															
$\frac{(T-b+1)}{n_{Tfb}}\sigma^2$	0.053	0.124	0.110	0.089	0.079	0.063	0.051								0.9573
	(0.001)	(0.002)	(0.003)	(0.006)	(0.010)	(0.015)	(0.032)								
$\frac{\sigma^2}{n_{Tfb}}$	0.056	0.133	0.110	0.095	0.071	0.069	0.052								0.9764
	(0.001)	(0.003)	(0.002)	(0.003)	(0.005)	(0.007)	(0.012)								
σ^2	0.059	0.111	0.110	0.079	0.088	0.049	0.115								0.8998
	(0.001)	(0.004)	(0.003)	(0.008)	(0.015)	(0.022)	(0.029)								
(4:72)															
$\frac{(T-f-23)}{n_{Tfb}}\sigma^2$	0.056							0.114	0.109	0.101	0.072	0.068	0.008	0.062	0.9534
	(0.001)							(0.003)	(0.003)	(0.005)	(0.009)	(0.013)	(0.022)	(0.059)	
$\frac{\sigma^2}{n_{Tfb}}$	0.059							0.119	0.110	0.104	0.073	0.071	0.009	0.065	0.9484
	(0.001)							(0.004)	(0.003)	(0.005)	(0.007)	(0.010)	(0.014)	(0.033)	
σ^2	0.059							0.106	0.095	0.087	0.074	0.085	0.031	0.063	0.8477
	(0.002)							(0.006)	(0.004)	(0.007)	(0.014)	(0.019)	(0.025)	(0.034)	

Note: Sample period = 1961–1970. Standard errors inside brackets. Estimation is made from monthly salaries in current prices by weighted least squares.

Table 4:10. *Estimated salary profiles of graduate engineers specializing in mechanical engineering (SMR)*

Model and variance specification	Annual increase in initial salary	Annual increase in active age interval							Annual increase in physical age interval								R
		0–4	5–9	10–14	15–19	20–24	25–29	30–34	–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	
(4:32)																	
$\frac{(T-b+1)}{n_{Tfb}}\sigma^2$	0.053 (0.001)	0.104 (0.002)	0.086 (0.002)	0.084 (0.004)	0.054 (0.007)	0.083 (0.010)	0.063 (0.015)	0.031 (0.026)	0.029 (0.002)	0.022 (0.002)	0.017 (0.004)	0.008 (0.006)	0.006 (0.008)	−0.044 (0.010)	0.059 (0.010)	−0.173 (0.040)	0.9814
$\frac{\sigma^2}{n_{Tfb}}$	0.054 (0.001)	0.107 (0.004)	0.090 (0.002)	0.088 (0.003)	0.054 (0.005)	0.089 (0.006)	0.058 (0.008)	0.078 (0.013)	0.032 (0.004)	0.019 (0.002)	0.016 (0.003)	0.009 (0.004)	−0.003 (0.005)	−0.022 (0.006)	−0.022 (0.010)	−0.058 (0.022)	0.9848
(4:41)																	
$\frac{(T-b+1)}{n_{Tfb}}\sigma^2$	0.051 (0.001)	0.126 (0.003)	0.108 (0.003)	0.095 (0.005)	0.059 (0.008)	0.078 (0.012)	0.042 (0.019)	0.045 (0.033)									0.9670
$\frac{\sigma^2}{n_{Tfb}}$	0.054 (0.001)	0.134 (0.003)	0.108 (0.002)	0.100 (0.004)	0.059 (0.005)	0.084 (0.007)	0.041 (0.009)	0.050 (0.014)									0.9771
(4:72)																	
$\frac{(T-f-23)}{n_{Tfb}}\sigma^2$	0.054 (0.001)								0.120 (0.003)	0.108 (0.003)	0.101 (0.005)	0.072 (0.007)	0.072 (0.010)	0.053 (0.013)	0.031 (0.023)	−0.010 (0.065)	0.9640
$\frac{\sigma^2}{n_{Tfb}}$	0.055 (0.001)								0.124 (0.005)	0.108 (0.003)	0.102 (0.004)	0.074 (0.006)	0.071 (0.007)	0.055 (0.009)	0.029 (0.014)	−0.000 (0.035)	0.9604

Note: Sample period = 1961–1970. Standard errors inside brackets. Estimation is made from monthly salaries in current prices by weighted least squares.

problem, the comparison already mentioned in chapter 2 has been made between the SAF and CF statistics for graduate engineers. It shows that those who belong to CF but have not been recorded by SAF have a higher (arithmetic) average salary in each age interval than those who are recorded in both lists. In the interval 50–59 years it is 27.5 % higher. The first group is also older than the second. The results from this investigation at least show that the estimates obtained from the SAF data cannot be used for generalizations for graduates in general and they do not even reveal a »typical» salary path for a graduate in SAF.

Model (4:72) is estimated from SAF data in current prices by weighted least squares regression. Table 4:11(A) gives the estimates for each of six educational groups when the residual variance is assumed to be proportional to physical age and inversely proportional to the number of employees, and table 4:11(B) gives the corresponding results when the variance only is inversely proportional to the number of employees.

The sensitivity to which variance specification is chosen is less than when the model was estimated from CF data. The two specifications almost produce the same fit. This is also true when the variance is assumed to be constant. These results are not, however, presented.

The estimates of the average increase in the initial salaries reveal a remarkable similarity between educational groups. They range between 6.1 % and 6.7 %. A general characteristic of the age dependent additional increases is that these increases are relatively low for very young employees. They reach a maximum at approximately the age of 30 years for university trained employees and a few years earlier for high school trained employees. These estimates thus show a better agreement with the curve TT in figure 4:1 than the estimates based on CF data. The salary increases of the oldest employees are even lower than for the youngest ones but this may be the result of the selection pointed out before.

A comparison between the educational groups reveals that the profiles of the three educational qualifications at university level engineering, business & economics and science are very similar. Science exhibits a minor deviation in the age class 30–34 years. The smaller average increase for science graduates may reflect a slower promotion because they are usually employed in fields which do not lead to top level jobs.

As a contrast the high school educational groups exhibit some dissimilarities. The increases obtained by those who have commercial education are in general higher than the increases obtained by engineers. They are even higher than the increases obtained by employees with a university degree which does not seem to be the case for engineers. Mincer's findings that the more the training, the higher the increases (see p. 84) do not therefore coincide with our findings for those with business education but do so for engineers.

In figures 4:4 and 4:5 estimated cohort profiles are drawn for employees who are assumed to be born in 1930. The first diagram permits comparisons in Swedish kronor between educational qualifications while the second diagram is drawn in logarithmic scale to make possible convenient comparisons of relative differences. It is assumed that those who are graduates of a university start work at the age of 25 and those who have high school certificates start work at the age of 20 years. The proper interpretation of the profiles is an illustration of the state of affairs during the sample period. The same conditions hardly prevail over such a long time that the profiles can be interpreted as estimates of profiles for those who were actually born in 1930. For the same reason they should not be used as predictions. The same careful interpretation should be applied to the illustrative calculations of lifetime salaries and rates of return to education in the following sections. All profiles illustrate salary paths in current prices as they developed during the sample period. They show no decrease in any age class, on the contrary, as we find from figure 4:4 the profiles are rather of the increasing exponential type. This diagram also clearly shows how salary differences in Swedish kronor increase by age, in particular between those who have a university degree and those who have not, The relative differences also increase a little. This is illustrated in figure 4:5 with the profiles of graduate engineers and high school engineers I. According to the diagram the increase takes place before the age of 40 years. After this age the relative difference remains approximately constant.

We now return to the three-dimensional cross section profile (4:38). Its shape is determined by the differences $(\gamma_i'-\beta)$ and by γ_i''. When any of these terms become negative the profile turns downwards. The estimates in tables 4:9–4:10 for engineers in SER and SMR show that the downturn in the active age dimension does not start until after 30 years in the labour market and in the physical age dimension after the age of 45 years.

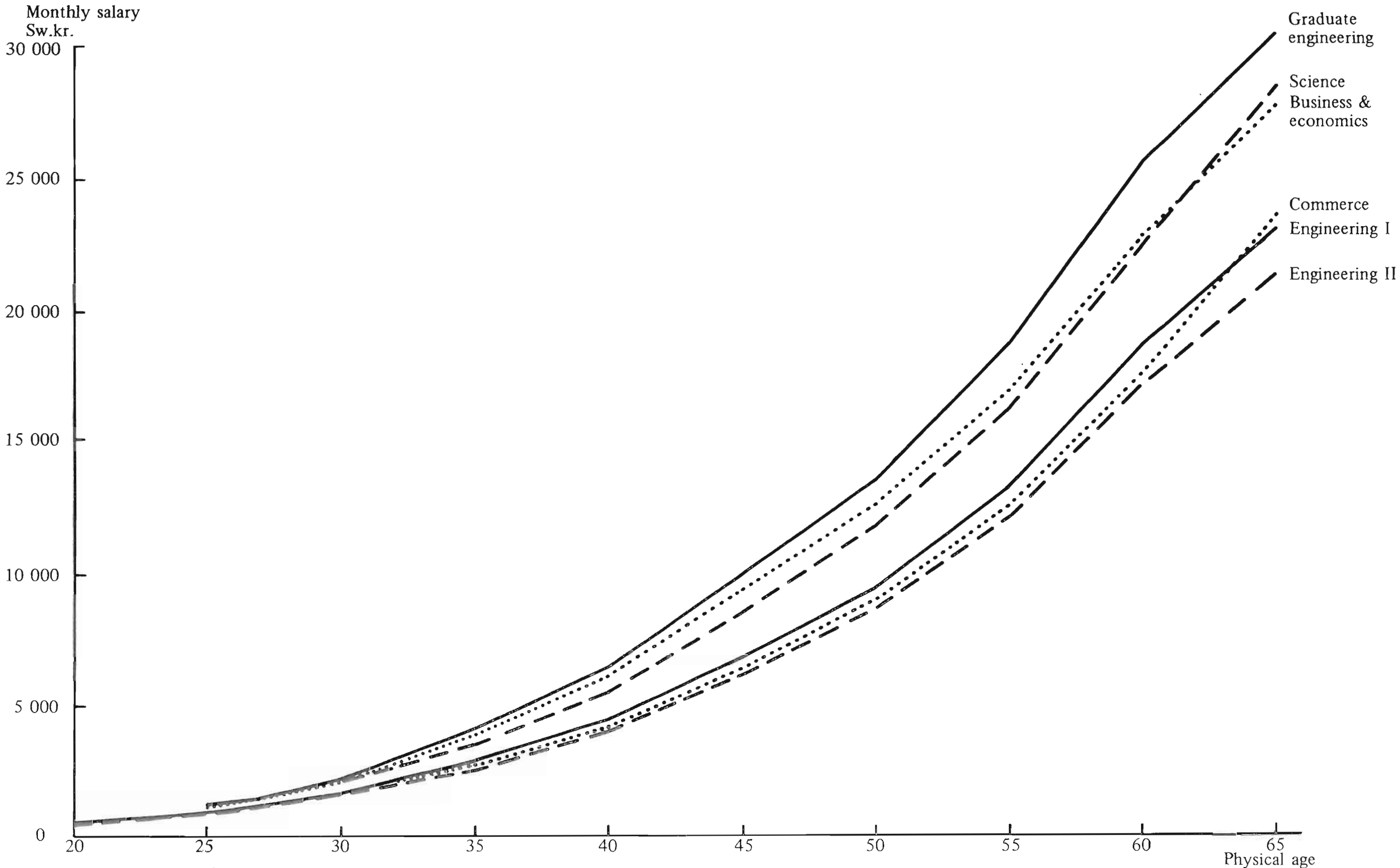
In the special case, when birth and labour market cohorts coincide, the cross section profile (4:43) does not turn downwards until the sum of salary increases due to active and physical age is less than the average annual shift. Even if the special case is not present, an analogous relation holds for a physical age-earnings profile. As the results in table 4:11 show, old employees do not obtain increases as high as the average shift in the cross sections. The break-even point is at about 50 years. It is important to note that the downturn of the profile does not necessarily imply that any employee receives a salary decrease. This is illustrated in figure 4:6 by estimated profiles for graduate engineers.

Table 4:11. *Estimated physical age-salary profiles from SAF data*

Educational qualifications	Annual increase in initial salary	Annual increase in physical age interval –21	22–25	26–29	30–34	35–44	45–59	60–	R	Sample period
A. *Residual variance proportional to physical age and inversely proportional to number of employees, specification (4:63)*										
University degrees										
Engineering	0.006		0.075	0.124	0.124	0.090	0.064	0.035	0.9966	1954–1969
	(0.001)		(0.009)	(0.003)	(0.002)	(0.002)	(0.002)	(0.011)		
Business & economics	0.065		0.069	0.133	0.122	0.087	0.059	0.040	0.9952	1952–1969
	(0.001)		(0.009)	(0.004)	(0.003)	(0.002)	(0.003)	(0.017)		
Science	0.063		0.090	0.124	0.102	0.090	0.066	0.051	0.9856	1956–1969
	(0.001)		(0.022)	(0.006)	(0.005)	(0.003)	(0.004)	(0.020)		
High school degrees										
Engineering I	0.064	0.081	0.106	0.115	0.103	0.085	0.066	0.041	0.9947	1952–1969
	(0.001)	(0.010)	(0.004)	(0.004)	(0.003)	(0.002)	(0.002)	(0.013)		
Engineering II	0.067	0.104	0.107	0.114	0.099	0.085	0.068	0.044	0.9966	1952–1969
	(0.000)	(0.010)	(0.003)	(0.003)	(0.002)	(0.001)	(0.002)	(0.012)		
Commerce	0.067	0.150	0.138	0.129	0.102	0.086	0.068	0.059	0.9955	1956–1969
	(0.001)	(0.008)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)	(0.014)		
B. *Residual variance inversely proportional to the number of employees, specification (4:64)*										
University degrees										
Engineering	0.065		0.069	0.125	0.123	0.089	0.063	0.034	0.9972	1954–1969
	(0.000)		(0.017)	(0.004)	(0.002)	(0.001)	(0.001)	(0.005)		
Business & economics	0.065		0.072	0.132	0.122	0.087	0.059	0.039	0.9944	1952–1969
	(0.001)		(0.021)	(0.005)	(0.003)	(0.002)	(0.002)	(0.009)		
Science	0.061		0.091	0.120	0.101	0.087	0.064	0.047	0.9870	1956–1969
	(0.001)		(0.038)	(0.008)	(0.004)	(0.002)	(0.002)	(0.010)		
High school degrees										
Engineering I	0.065	0.080	0.108	0.115	0.104	0.086	0.067	0.042	0.9969	1952–1969
	(0.000)	(0.021)	(0.005)	(0.003)	(0.002)	(0.001)	(0.001)	(0.005)		
Engineering II	0.067	0.099	0.109	0.112	0.100	0.085	0.068	0.044	0.9976	1952–1969
	(0.000)	(0.024)	(0.005)	(0.003)	(0.002)	(0.001)	(0.001)	(0.006)		
Commerce	0.066	0.140	0.141	0.124	0.102	0.084	0.067	0.058	0.9948	1956–1969
	(0.001)	(0.018)	(0.006)	(0.004)	(0.003)	(0.002)	(0.002)	(0.007)		

Note: Standard errors inside brackets. Weighted least squares regression applied to model (4:72) and with data in current prices.

Figure 4:4. *Estimated cohort profiles by education for employees assumed to be born in 1930*



Note: The profiles are obtained from the estimates in table 4:11.

Figure 4:5. *Estimated cohort profiles for graduate engineers and high school engineers assumed to be born in 1930*

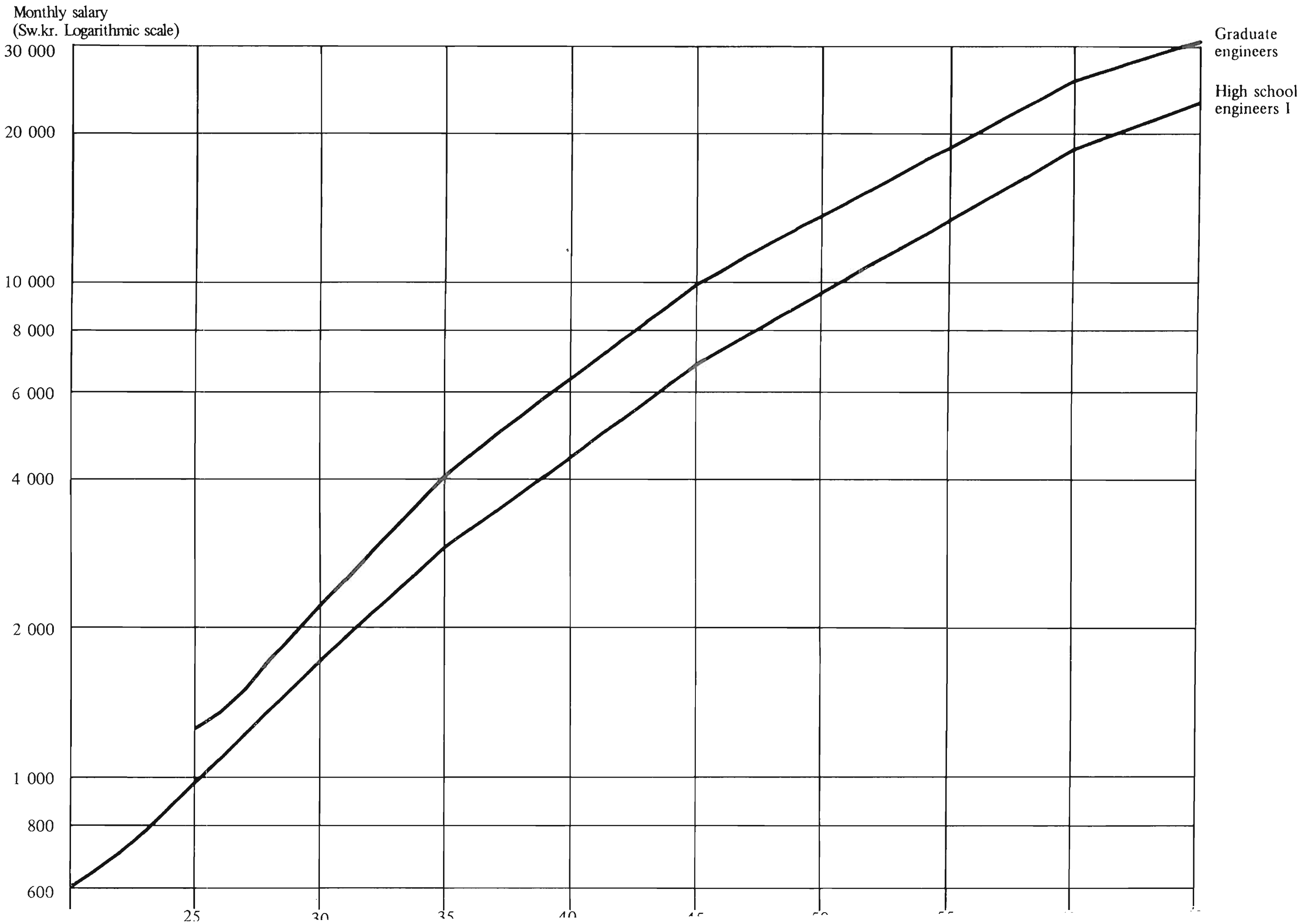
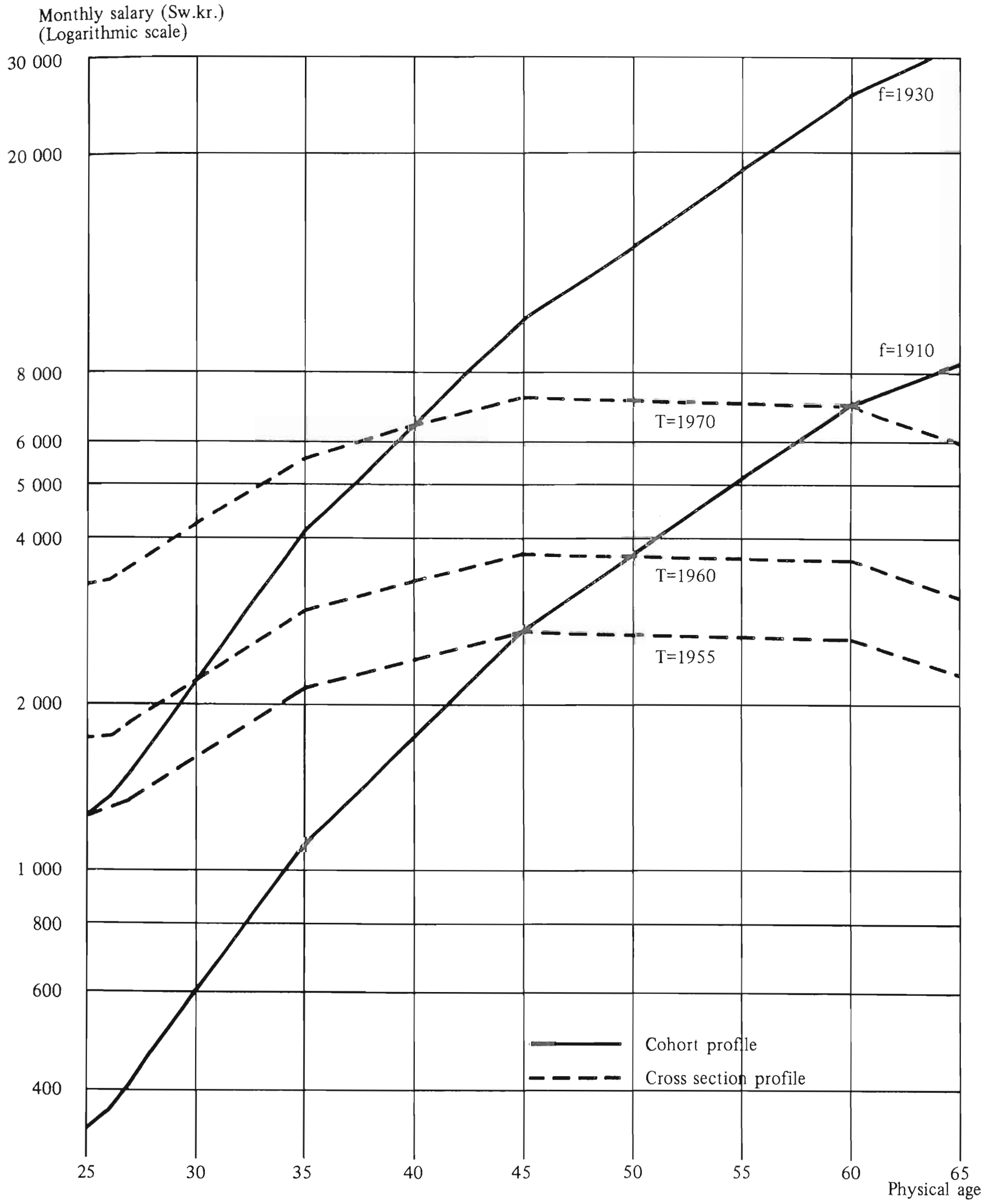


Figure 4:6. *Estimated cohort and cross section profiles for graduate engineers*



Note: The profiles are obtained from the estimates in table 4:11.

4.3.7 *Some remarks on possible refinements*

The formal representation in section 4.2 shows that increases in earnings could in principle vary by calendar time. In section 4.3 the dependence on active and physical age has been stressed and the dependence on calendar time has been neglected. None of the models suggested represents effects of supply and demand and of negotiations, and these factors may have effects associated more with calendar time than with age. No attempt will be made at this stage to explain this in detail, but a few remarks may be of some interest.

The supply of educated labour in Sweden has increased during the whole sample period, but the greatest increase has taken place over the last few years (see chapter 3). While there are no indications of short-run fluctuations of supply, demand shows pronounced fluctuations. The effect on earnings of both supply and demand should be most observable for young employees since they show a higher mobility and since the salaries of new employees are not regulated in negotiations. When demand is high relative to supply, earnings of young employees would increase more than the earnings of middle aged and old ones. The age-earnings profiles would then become flatter. The opposite should be true in the case of excess supply. Table 4:12 shows the number of jobs available for engineers and technicians advertised in Dagens Nyheter, which is a measure of demand for this kind of labour. There are peaks in 1960, 1965 and 1969 and troughs in 1958, 1962 and 1967, all in close agreement with the general business cycle in Sweden. A comparison with the residuals from the estimated model (4:72) for graduate engineers and high school engineers I in table 4:13 does not reveal any systematic positive association between demand for labour and salaries, not even for young employees. This preliminary study does not thus give support to the theory of a short-run sensitivity of earnings to changes in demand. There is, however, an indication of a systematic pattern in the residuals. Negative residuals are most frequent before 1961 and during the last years of the sample period, while predominantly positive residuals are obtained for the first half of the sixties. This may be explained by a relatively low increase in productivity in Swedish industry in the fifties and a relatively high increase in the sixties. The explanation behind the small or negative residuals during the last years of the sample period is probably the increased supply. These conclusions are very tentative and in a more refined analysis the effects of other factors like price increases and negotiations should also be investigated at the same time.

To link such an analysis to the model used in this study would permit an explanation of both long-run and short-run changes in the profiles, which could be useful for predictions and for calculations of lifetime earnings.

Table 4:12. *Selected labour market indicators in 1954–1969*

Year	Number of jobs for engineers and technicians advertised in Dagens Nyheter* (thousands)	Percentage increase in productivity**	Percentage increase in consumer price index	Percentage increase in negotiated salary***
1954	12.1	6.4	0.8	0.0
1955	12.1	1.9	3.1	9.1
1956	11.6	5.5	4.5	4.0
1957	12.4	4.8	4.3	3.7
1958	11.2	4.7	4.8	3.5
1959	15.2	6.8	0.7	1.9
1960	18.1	5.0	3.9	3.8
1961	16.9	4.1	2.5	3.5
1962	15.0	5.4	4.3	4.7
1963	14.5	7.0	2.9	3.5
1964	15.9	8.0	3.4	4.0
1965	17.7	7.5	5.0	3.3
1966	12.8	5.0	6.3	7.2
1967	8.3	8.1	4.5	5.4
1968	9.3	8.8	1.9	5.0
1969	14.6	9.0	2.8	3.5

* Figures for 1954–1963 are estimates.

** Value added in constant prices per manhour.

*** Men only.

Sources: Unpublished sources from Dagens Nyheter, The Industrial Institute for Economic and Social Research and the Swedish Employers' Confederation. Consumer price index is published in the Statistical year book for Sweden.

Table 4:13. *Logarithmic residuals from estimated physical age-earnings profiles by education, physical age and time in 1954–1969*

Year	Graduate engineers			High school engineers I			University degree in business & economics			High school certificate in commerce		
	25	42.5	55	21	37.5	55	25	42.5	55	21	37.5	55
1954	−0.032	0.020	−0.074	−0.072	−0.030	−0.039	−0.062	0.059	0.015	–	–	–
1955	−0.054	0.043	−0.010	−0.038	0.015	0.003	−0.063	0.024	−0.008	–	–	–
1956	−0.039	0.031	−0.025	0.015	−0.004	−0.005	−0.027	0.056	−0.032	0.034	−0.020	0.082
1957	0.001	0.017	−0.041	0.055	−0.010	−0.005	−0.057	0.075	−0.097	−0.018	0.000	0.004
1958	−0.027	0.008	−0.023	0.040	−0.028	−0.024	−0.052	−0.023	−0.073	0.015	−0.006	−0.022
1959	−0.052	−0.010	−0.061	0.021	−0.051	−0.049	−0.009	−0.047	−0.106	−0.003	−0.027	−0.016
1960	−0.032	−0.024	−0.046	0.044	−0.047	−0.041	−0.094	−0.035	−0.088	−0.013	−0.035	−0.015
1961	0.027	0.019	0.034	0.078	0.011	0.007	0.001	0.014	−0.011	0.041	0.013	0.014
1962	0.049	0.019	0.050	0.068	0.028	0.028	0.053	0.012	0.026	0.047	0.016	0.050
1963	0.052	0.020	0.037	0.073	0.021	0.027	0.031	−0.013	0.057	0.039	0.003	0.027
1964	0.043	0.009	0.030	0.041	0.016	0.008	0.016	−0.029	0.085	0.074	−0.013	0.023
1965	0.057	0.006	0.035	0.015	0.024	0.003	0.037	−0.020	0.060	0.040	0.008	0.002
1966	0.040	0.002	0.035	0.004	0.036	0.023	0.023	−0.001	0.047	0.087	0.025	0.012
1967	0.026	−0.002	0.030	−0.007	0.022	0.009	0.018	0.010	0.024	0.051	0.035	0.013
1968	−0.020	−0.028	0.018	−0.097	0.008	0.007	0.008	0.009	0.018	−0.014	0.022	0.007
1969	−0.074	−0.055	−0.007	−0.128	−0.026	−0.008	−0.040	−0.015	−0.023	−0.016	−0.039	−0.076

Note: The residuals are obtained in the estimation of model (4:72) from SAF data. The parameter estimates have been presented in table 4:11(B).

4.4 LIFETIME SALARIES AND RATES OF RETURN TO EDUCATION

The lifetime salary of an employee is the sum of all the salaries he has earned during his active time on the labour market. The expected lifetime salary of a group of individuals with a common characteristic, for instance the same education, can be estimated by a sample mean of lifetime salaries, but usually it is not possible to follow the salary flow of identical individuals and second best methods have to be used. One method is to construct a hypothetical promotion path from jobs with a low ranking to jobs with a high ranking and to use the current salaries at each job level (see for instance Bailey & Schotta [1969], SACO [1968]). Frequently lifetime salaries (earnings) are calculated as the area underneath a cross section profile although this profile does not necessarily have the same shape as profiles of identical individuals.★ Examples of studies when this method has been used can be found in SACO [1968].

A better method is perhaps to estimate cohort profiles by model (4:32) or if this is not possible, by model (4:72), and then to calculate the sum of the estimated salaries as defined in (4:15) or (4:16). If the estimates are obtained as the antilogarithms of unbiased estimates of logarithmic average salaries, these estimates are not themselves unbiased. If, for instance, the logarithmic residuals of the model are normally distributed the procedure described gives unbiased estimates of median salaries and not of geometrical average salaries. The estimated lifetime salary would then be a sum of estimated median salaries, (to be distinguished from the median lifetime salary).

Provided an exponential trend can be fitted to the consumer price index (see p.110) a lifetime salary calculated from a model, which is estimated from data in constant prices, can be seen as a lifetime salary obtained from a model for data in current prices discounted by the average price increase. Suppose (4:32) is discounted to the base year b by the discount factor δ. The discounted salary at T is then

$$L^D_{Tfb.} = L_{Tfb.} \exp[-(T-b)\delta] = \exp[\alpha + \beta b + \sum_{i=1}^{c} (\gamma_i' - \delta) D_i' + \sum_{i=0}^{c} \gamma_i'' D_i'' + \epsilon_{fb} + \sum_{t=b+1}^{T} \vartheta_{tfb}]; \qquad (4:83)$$

If for a particular b the base of the price index P_T in (4:81) is changed to 1 for T=b, a comparison between (4:83) and (4:82) shows that except for the price residual ζ_T the two expressions are identical for $\delta=\pi$. As the observed deviations

★ Siegfried [1971] following Becker [1964] used an exponential function $(1.0125)^{t-b}$ to adjust cross sections profiles of Ph.D. economists to cohort profiles.

from the price trend need not be perfectly uncorrelated with the independent variables in the model, this relation between real and nominal lifetime salaries is only approximately valid. If there is a systematic trend in the price *increases* and the exponential trend does not give a good approximation, real lifetime salaries cannot be calculated satisfactorily from profiles estimated from data in current prices. The reason is that the first salaries in a career, which would be erroneously deflated by the average price increase, are very important in the discounting procedure.

Table 4:14 contains estimates of lifetime salaries for employees assumed born in 1930. The calculations have been made from profiles estimated in current prices. It is assumed that the first salary was earned at the age of 25 for employees with a university degree and at 20 for those who have a high school certificate. Three alternative discount rates are used, 5 %, 10 % and 15 %. All lifetime salaries are discounted to the age of 20 years. The choice of discount rate greatly affects the lifetime salaries in Swedish kronor as well as the relative differences as is shown by the index numbers in the table. To compare lifetime salaries is a method of making comparisons which allows for differences in active time in the labour market. At a discount rate of 5 % university trained employees obtain between 18 % and 44 % higher a remuneration than high school trained employees but at the rate of 15 % the relative differences only range from 2 %

Table 4:14. *Estimated lifetime salaries by education*

Educational qualifications	Lifetime salaries (Sw.kr. 1000)						Active time
	Discount rate 5%	Index	Discount rate 10%	Index	Discount rate 15%	Index	
University degrees							
Engineering	1370	140	421	129	173	111	25–65
Business & economics	1271	130	396	122	164	105	25–65
Science	1213	124	378	116	158	102	25–65
High school degrees							
Engineering I	1028	105	348	107	170	109	20–65
Engineering II	943	97	320	98	158	101	20–65
Commerce	977	100	326	100	156	100	20–65

Note: The calculations are made on the assumption that the year of birth is 1930 and that employees with high school education start work at the age of 20 while employees with university education do not start until they are 25 years old. All salaries are discounted to the age of 20 years. They are obtained from cohort profiles calculated from the estimates in table 4:11(B). The lifetime salaries are not independent of the price increases during the sample period.

to 11 % (see the index numbers in table 4:14). The choice of cohort may influence the result of the comparison. If two educational groups have the same average shift β in the profiles the choice of cohort has no influence at all, but the larger the difference in β, the greater the effect the cohort has on the comparison. The estimates in table 4:11 show max-differences of 0.006 which means that the ratio of two lifetime salaries changes by not more than 6 % over 10 years.

The application of cross section profiles for calculations of lifetime salaries was criticized above. Would the lifetime salary obtained from a cross section profile differ very much from the salary obtained from a cohort profile? Assuming that the model (4:32) is true the answer depends on β and the discount rate used. Suppose the present value of the salary stream is discounted to the first year in the labour market, say $b=T_0$, by the discount rate $(\exp[\delta]-1)$. (4:83) gives the discounted profile. We now find that the cross section profile (4:38) and the discounted profile (4:83) are identical for $b=T_0$, provided that $\beta=\delta$ or, which is the same thing, the average increase in initial salaries equals the discount rate. Under the assumptions made this result can be generalized to the following. A lifetime salary calculated from a cohort profile at a certain discount rate can also be obtained from a cross section profile if the discount rate is reduced by the average increase in initial salaries (average shift in the cross sections).

An alternative way of comparing the remuneration associated with two educational qualifications is to calculate relative rates of return to education. For instance, what return does a high school engineer obtain if he enters college to graduate as an engineer? There are several ways of calculating a rate of return which answers the question. In principle we should find a rate which equals the sums of two discounted *net* income streams. Net means that all incomes including scholarships and grants are added and all costs associated with the earning of these incomes including for instance tuition are deducted. To obtain a rate of return which is relevant for an individual choice whether or not to obtain additional education the calculations should be done after tax. The result is usually called a private rate of return for education, to be distinguished from a social rate of return which reflects social incomes and costs of education.

The calculations of relative private rates of return presented in table 4:15 only serve as illustrations and are based on some very crude assumptions. Costs associated with a college training are assumed to be equal to scholarships and grants received. Cost of living is assumed to be the same independent of education. The income streams are projections based on the estimates in table 4:11(B) and the average price increase during the sample period is thus wowen into the lifetime salaries. As an approximation the real rate of return may be obtained

Table 4:15. *Relative rates of return to education*

Educational qualifications	Rate of return (%)
Degree in engineering/certificate in engineering I	17
Degree in business & economics/certificate in commerce	18
Degree in business & economics/certificate in engineering I	15
Degree in science/certificate in engineering I	14

Note: The rate of return r is determined as the value of r for which

$$\sum_{t=0}^{45} \frac{Y(t) - X(t)}{(1+r)^t} = 0;$$

where Y(t) and X(t) are nominal income streams before tax projected by the model (4:72). The birth year is assumed to be 1930 for both groups.

after deduction of the average price increase of 4 %. The rates of return are as a matter of fact *not* calculated net of taxes. They should therefore be interpreted as measures of the gross salary remuneration of one education relative to another standardized for differences in active time. To interpret the numbers in table 4:15 as estimates of actual relative rates of return for those who were born in 1930 would hardly be meaningful. To do this a careful evaluation and forecast of the factors which determine the salary increases in the future would be necessary. The rates of return now calculated are primarily descriptions of the state of affairs during the sample period. With this interpretation in mind we find that the rate of return for a degree in engineering is 17 % relative to a high school certificate in engineering. The return to college training in economics or science is a few percentage units less. These results approximately agree with results reported from other studies. In Blaug, Peston & Ziderman [1967] (table 10, p. 79), the estimated private return to an ordinary degree relative to a Higher National Certificate (H.N.C.) is 9 %. This estimate, however, is net of income-tax. A rate of return before tax would probably be a few percentage units higher. If furthermore 4 percentage units are subtracted from the estimates in table 4:15 to allow for price increases in Sweden the results from the two studies are approximately the same. In Fase [1969] (table 25 p. 81), the rate of return to a degree in engineering relative to secondary education (II) is estimated to 12.6 % after tax and 15.1 % before tax. Both estimates are obtained from cross section data. They are thus net of economic growth and price increases. If 6–7 percentage units are subtracted from the estimates in table 4:15 they become approximately comparable with the Dutch estimates. The result 10–11 % is a little less than Fase obtained.

4.5 INTRODUCTION TO A CROSS SECTIONAL ANALYSIS OF EARNINGS

The models obtained so far have been used to analyse earnings profiles of employees differentiated by education. No other concepts have been used to define homogeneous groups of employees. Increases in earnings have been analysed as a function of age. Except for a few references to the human capital literature and a digression on short-run variations in salary increases, the mechanism behind the increases by age has not been analysed.

In the following chapter factors other than age and education will be brought into the analysis to explain salary differences in cross sections. It will be shown that the job level is strongly differentiatory and that promotion from one level to another is thus one important component of salary increases. Other factors which may also differentiate between salaries are for instance job family, industry and cost of living area. These factors do not perhaps play the same instrumental role as job level in explaining salary increases, they may rather be used to define more or less homogeneous groups of employees in the same way as the factor education has been used. However, this does not of course exclude the possibility that an employee may obtain a salary rise by for instance moving from a low-paying industry to a high-paying one.

In this section the model of a cross section profile (4:38) will be modified, first by the introduction of salary increases due to the factor job level and secondly by combining in one model cross section profiles of several groups of employees with the *same* salary increases. Finally these two approaches will be brought together in one model of the salary structure in a cross section. All these exercises serve the purpose of an introduction to the models used later in the analysis of the salary structure in Swedish industry.

In the cross section model (4:38) $(\gamma_i'-\beta)$ may be interpreted as a salary difference due to one year's difference in experience. As information about active age is not available in a cross section of SAF data the model (4:38) cannot be applied unless active age is replaced by some other variable. In the human capital theory active age is usually a proxy for experience. Experience, theoretical or practical, is associated with the skills and responsibility required for a job, i.e. the job level. The higher the job level, the more the experience required and the more experience an employee gains, the greater his chances are of promotion. As a substitute for active age, it should thus be possible to measure experience by the job level attained. In doing so one has to disregard the increase in experience and also in salary which an employee gains at a given job level and approximate the range of salaries at this level by its midpoint.

If $(\gamma_i'-\beta)$ is now interpreted as a salary increase due to increased experience, γ_i'' may be interpreted as a salary increase which is mainly associated with efficiency in work for a given level of experience, i.e. at a given job level.★ Young employees are supposed to be more efficient than middle aged and old ones. The rate of salary increase due to promotion (increased experience) may in principle depend on the age at which promotion is obtained. Interactions between job level and age will be investigated in chapter 5 but to keep the exposition in this section simple no interactions are introduced at this stage. Under these assumptions and under the additional assumption that at time T the p-th job level is attained of maximum m levels, a first step towards a reformulation of the cross section model gives, when the stochastic components are omitted

$$E(\ln L_{Tfb.}) = \alpha + \beta T + \sum_{j=2}^{p} \lambda_j + \sum_{i=0}^{c} \gamma_i'' D_i'' ; \tag{4:84}$$

where λ_j is the effect of a promotion from job level j−1 to j.

Although there is certainly a relation of the nature described above between job level and active age too much importance should not be attached to it. Promotion from one job level to another probably also contains other aspects which cannot be measured by active age, and vice versa. A less restrictive approach is to say that one important explanation for the salary increases associated with physical age, when active age is omitted, is promotion. Its contribution to the increases can be measured if the factor job level is introduced. The remaining effects associated with age are then picked up by the age factor.

In a second step of reformulation a dummy variable Y_j is introduced for each job level. Y_j takes the value 1 for observations at the j-th job level, otherwise its value is 0. A new parameter ω_j is also defined; $\omega_j = \sum_{r=2}^{j} \lambda_r$; j=2,..., m. To simplify the analysis further and to treat the two factors alike the partial relation between salary and physical age in (4:84) is approximated by a new function. In (4:84) salaries increase by a constant percentage inside each age interval. We now choose to represent the salaries in each interval by one value only. Instead of a polygon we thus use a step function. The new model is now

$$E(\ln L_{Tfb.}) = \mu' + \sum_{j=2}^{m} \omega_j Y_j + \sum_{i=1}^{c} \kappa_i Z_i ; \tag{4:85}$$

μ' is the expected logarithmic salary of those who belong to the first age interval and to the first job level. Z_i is a dummy variable which takes the value 1

★ See the Swedish terminology »duglighet» and »prestation» in SAF–SIF [1968].

for observations in the i-th age interval and is otherwise 0. κ_i shows how much higher the salary is in age interval i, i=1, ... , c; compared to the first interval (i.e. interval 0) independently of job level and ω_j shows how much the salary at job level j, j=2, ... ,m; deviates from the salary at the first level.

We will now turn to the treatment of more than one group of employees. A group may be defined by education, industry, cost of living area and so forth. Suppose there are g groups of employees who experience the same salary increases but have different initial salaries. Following expression (4:38) the salary structure of the g groups can then be written

$$E(\ln L_{Tfb.}) = \sum_{h=1}^{g} (\alpha_h + \beta T) X_h + \sum_{i=1}^{c} (\gamma_i' - \beta) D_i' + \sum_{i=0}^{c} \gamma_i'' D_i''; \qquad (4:86)$$

where the X_h 's are dummy variables. The g intercepts can be reformulated as deviations η_h from a common intercept μ''.

$$E(\ln L_{Tfb.}) = \mu'' + \sum_{h=1}^{g} \eta_h X_h + \sum_{i=1}^{c} (\gamma_i' - \beta) D_i' + \sum_{i=0}^{c} \gamma_i'' D_i''; \qquad (4:87)$$

μ'' and the g η_h's are, however, not uniquely defined as these six parameters correspond to only five »composite» parameters $(\alpha_h + \beta T)$. Unique parameters can be obtained if an appropriate linear constraint is imposed on μ'' and the η_h's. A constraint of the following type is usually chosen

$$\sum_{h=1}^{g} W_h^{\eta} \eta_h = 0; \qquad (4:88)$$

where the W_h^{η} 's are suitably chosen weights the sum of which is one. From (4:86) and (4:87) it follows that

$$(\alpha_h + \beta T) = \mu'' + \eta_h; \quad h=1, \ldots ,g \qquad (4:89)$$

Solving (4:88) and (4:89) for μ'' and η_h gives

$$\begin{cases} \mu'' = \Sigma W_h^{\eta} (\alpha_h + \beta T) = \Sigma W_h^{\eta} \alpha_h + \beta T; & (4:90a) \\ \eta_h = (\alpha_h + \beta T) - \mu'' = \alpha_h - \Sigma W_h^{\eta} \alpha_h; \quad h=1, \ldots ,g & (4:90b) \end{cases}$$

μ'' is thus a weighted average of the initial salaries and η_h is the deviation of the initial salary of group h from this average. The weights obviously determine the interpretation of μ'' and η_h. This problem will be dealt with in some detail in chapter 5.

The characteristic features of the two expressions (4:85) and (4:87) give a new model which will be used in a more or less generalized form in chapter 5 to analyse the salary structure in cross sections of Swedish industry.

$$\begin{cases} E(\ln L_{Tfb\cdot}) = \mu + \sum_{h=1}^{g} \eta_h X_h + \sum_{j=1}^{m} \omega_j Y_j + \sum_{i=0}^{c} \kappa_i Z_i \; ; & (4{:}91a) \\ \sum_{h=1}^{g} W_h^{\eta} \eta_h = 0; & (4{:}91b) \\ \sum_{j=1}^{m} W_j^{\omega} \omega_j = 0 \; ; & (4{:}91c) \\ \sum_{i=0}^{c} W_i^{\kappa} \kappa_i = 0 \; ; & (4{:}91d) \end{cases}$$

A particular set of weights was chosen in (4:85), namely $W^{\omega} = W_0^{\kappa} = 1$ and all other weights equalled zero. These weights gave a convenient interpretation of the parameters ω_j and κ_i, but they are not the only feasible weights. To keep the model general, no specific weights are specified in (4:91). We will return to this problem in chapter 5. μ is interpreted as an overall average expected logarithmic salary and the parameters η_h, ω_j and κ_i express deviations from this average due to group, age and job level respectively. Models of this kind have previously been applied by for instance Hill [1959], Klevmarken [1968a] and Holm [1970].

CHAPTER 5

A CROSS SECTIONAL ANALYSIS OF SALARY DIFFERENCES

In the previous chapter a model was developed for the analysis of both cohort and cross section age-earnings profiles, and the last section of the chapter indicated how this model could be transformed so as to cover a more detailed cross section analysis. Salary differences in several cross sections from the SAF salary statistics will be analysed in this chapter by models of this kind, i.e. models of the class of general linear models of less than full rank. The property not of full rank implies that the parameters of a model are not all unique, and that the normal equations do not have a unique solution. In the traditional application of the general linear model, which is to investigate estimable linear functions of the parameters, any solution of the normal equations will do to obtain a least squares estimate of the linear function or to make some of the usual tests. With this application in mind it is easy to understand that it is the multiplicity of solutions to the normal equations which is usually seen as the main problem. For instance, Searle writes (Searle [1971] p. 209): »The source of difficulties with the model not of full rank is that the normal equations ... have no unique solution.» The problem is thus primarily seen as an estimation problem, one of finding an arbitrary solution to the normal equations. This can be done by imposing appropriate constraints on the solution or by utilizing the approach based on generalized inverses of a matrix. The two are of course intimately related.

The present application differs somewhat from the traditional one. Our interest is not focused on estimable linear functions of parameters, rather we look for a convenient and general description of the salary structure. It should be possible to use it for many different comparisons including those which involve non-estimable linear parameter functions. The fundamental problem is therefore the non-uniqueness of the parameters and their interpretation. It can formally be solved by an appropriate standardization of the parameters, i.e. by imposing constraints on the parameters in order to obtain uniqueness. The constraints cannot be chosen arbitrarily. They should not reduce the rank of the model unless this is desired for non-statistical a priori reasons. We may then speak of restricted models.* However, there are still an infinite number of possible ways of standardization, i.e. an infinite number of feasible constraints. Each set of constraints implies a certain interpretation of the parameters. In one application one interpretation may be more convenient than another, while in another application another interpretation may be preferred. In the more traditional approach the interpretation of the parameters implied by the constraints used is irrelevant, but this is not so in the present approach. The result of comparisons sometimes depends on the choice of constraints.

It is convenient to change somewhat the notation used in chapter 4 and to state the basic features of linear models not of full rank using this new notation before the empirical analysis is started. This is done in section 5.1. The choice of constraints and the interpretation of the parameters implied by the constraints are also treated in the same section. The first part of section 5.2 is a survey of the variables and particular models used in the analysis of each cross section and the second part contains the empirical results. The empirical analysis in this chapter differs from the previous one by the fact that individual salaries are used rather than averages.

5.1 MODELS AND METHODS

5.1.1 *Basic features of the models*

Suppose individual salaries of a particular group of employees are observed in a given month. The problem given is to analyse how these salaries depend on factors such as education and job. In order that the analysis may be rendered more simple, it will be limited for the moment to these two factors.

---

* Searle [1971] p. 205 suggests the following terminology:
» ..... Sometimes, however, more explicit definitions inherent in the model result in relationships (or restrictions) existing among the parameters of the model. These are considered part and parcel of the model. For example, the situation may be such that the parameters of the model satisfy the relation $\alpha_1 + \alpha_2 + \alpha_3 = 0$; that is, we take this not as a hypothesis to be tested but as a fact, without question. Relationships of this nature, existing as an integral part of a model, will be called *restrictions on the model.* Their origin and concept is not the same as that of relationships that sometimes get imposed on the solutions of normal equations in order to simplify obtaining those solutions; those relationships will be called *constraints on the solutions.* »

In the present study it is preferred to use the terms 'constraints' and 'restrictions' in the more technical meaning indicated above in the text.

The following notation is introduced:

p = number of educations

q = number of jobs

n_{ij} = number of employees with education i and job j; $i=1,...,p$. $j=1,...,q$.

$n_{i\cdot} = \sum_{j=1}^{q} n_{ij}$ = number of employees with education i; $i=1,...,p$.

$n_{\cdot j} = \sum_{i=1}^{p} n_{ij}$ = number of employees with job j; $j=1,...,q$

$n_{\cdot\cdot} = \sum_{i=1}^{p} n_{i\cdot} = \sum_{j=1}^{q} n_{\cdot j} = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}$ = the total number of employees, i.e. the total number of observations.

L_{ijk} = the salary of the k-th employee with education i and job j. $k=1,..., n_{ij}$.

Suppose now that L_{ijk} can be looked upon as the product of two components, one which depends only on education and job and one which does not depend on either of these two factors.

$$L_{ijk} = \phi_{ij} \times u_{ijk}; \qquad (5:1)$$

In the following sections ϕ_{ij} will be treated as a non-stochastic component and u_{ijk} as a stochastic variable. What is the nature of u_{ijk}? If the observations are obtained by a sampling survey u_{ijk} may simply be looked upon as a sampling error. But in econometrics all possible individuals are very often surveyed which is also the case in this study, except for non-response, and the usual sampling error interpretation is not possible. u_{ijk} is rather a part of the salary setting process, i.e. this process is stochastic and not primarily generated by the survey methods. This of course does not exclude the possibility that there are measurement errors and the like which may also be included in u_{ijk}.

It is very common in econometrics to look upon u_{ijk} as a residual which contains everything not explained by ϕ_{ij}. This view does not necessarily conflict with the interpretation of u_{ijk} given above. On the contrary a list of factors possibly omitted from the non-stochastic component may help to specify realistic properties of u_{ijk}.

The stochastic properties given to u_{ijk} explicitly or implicitly imply that u_{ijk} follows a certain distribution. The observations are then considered as realizations drawn from this distribution. What inferences can then be made from this kind of model? The answer depends on how general the model is believed to be. It has already been mentioned that the whole population of individuals for a given time period has been surveyed. Does the model also have validity for other populations of individuals at other time periods? Usually the econometrician wants to give his model some validity outside the sample period or outside the surveyed population of individuals, but on the other hand he readily admits that the model is probably no good »far from» the sample period and for populations »different» from the surveyed one. In econometrics this is named »structural changes». Gadd and Wold [1964] have even developed a measure of structural changes, the Janus coefficient. When econometric models are used in forecasting it is common for anticipated structural changes during the forecasting period to be built into the model after the estimation is completed. It is usually difficult to specify how valid a model is outside the population and time period observed; to be strictly formal it is only valid inside this domain. Is it then at all interesting to make inferences for a hypothetical distribution valid for a limited group of individuals and for one historical time period, when we know the actual realizations of the process? If the analysis is to have any value besides a description of historical facts, we have to admit that the results may reveal something of an economic structure, which is valid in a more general context. An important observation is, however, that an inference drawn to a domain outside the observed population of individuals and sample period is usually not a statistical inference but an inference based on the economists expert judgement. This may perhaps be seen as a result of insufficient model building and a lack of confidence in econometric models.

The purpose of bringing these questions up has not been to analyse philosophical aspects of econometric model building, but rather to indicate how the author looks upon the statistical methods in this context and how, in his opinion, the results of this study should be interpreted.

Following the procedures of the two preceding chapters, (5:1) is converted into logarithms.

$$\ln L_{ijk} = \ln\phi_{ij} + \ln u_{ijk}; \tag{5:2}$$

The specific properties of u_{ijk} will not be specified until needed, except that

$$E(\ln u_{ijk}) = 0; \tag{5:3}$$

The expected value of $\ln L_{ijk}$ is then $\ln\phi_{ij}$. In analogy with the models in chapter 4, $\ln\phi_{ij}$ is broken down into three components

$$\ln\phi_{ij} = \mu + \alpha_i + \beta_j; \qquad i=1,\ldots,p. \quad j=1,\ldots,q \tag{5:4}$$

μ is the expected average salary and α_i and β_j the deviations from the average due to education i and job j respectively. To simplify the notation ϵ_{ijk} is substituted for $\ln u_{ijk}$ and ℓ_{ijk} for $\ln L_{ijk}$. We now obtain

$$\ell_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}; \quad i=1,\dots,p. \quad j=1,\dots,q \tag{5:5}$$

$$E(\epsilon_{ijk}) = 0; \tag{5:6}$$

To obtain (5:5) in a form similar to the usual linear regression model it can be rewritten once again with dummy variables, one for each education and one for each job.

$$\ell_{ijk} = \mu + \sum_{i=1}^{p} \alpha_i X_i + \sum_{j=1}^{q} \beta_j Y_j + \epsilon_{ijk}; \tag{5:7}$$

The value of X_i is unity when an employee has education i and is otherwise zero, and the value of Y_j is unity when he belongs to job j and is otherwise zero.

For each set of dummy variables corresponding to one of the factors education and job and for each observation, it is true that the sum of the dummy variables is always equal to one,

$$\sum_{i=1}^{p} X_i = \sum_{j=1}^{q} Y_j = 1; \tag{5:8}$$

i.e. the dummy variables of a set are linearly dependent. The parameters of the model (5:7) are then not unique. To see this, add an arbitrary constant to all α_i (or all β_j). It is then possible to compensate for the increased α_i's (β_j's) by subtracting the same constant from μ, without changing the value of ℓ_{ijk}. To obtain unique parameters we choose to impose constraints. As there are two independent linear dependences among the dummy variables, two independent linear constraints are needed. They are traditionally of the following form.

$$\sum_{i=1}^{p} W_{i.}\alpha_i = 0; \tag{5:9}$$

$$\sum_{j=1}^{q} W_{.j}\beta_j = 0; \tag{5:10}$$

The weights $W_{i.}$ and $W_{.j}$ have to be chosen according to certain rules which will be discussed in a following section.

As the dependent variable is the natural logarithm of monthly salaries, $100\alpha_i$ can approximately be interpreted as the percentage deviation from the average salary due to education i, and $100\beta_j$ analogously. In the present applications the approximation is rather good in the interval −15 % to +15 %. Similarly for two educational qualifications i and r $100(\alpha_i - \alpha_r)$ is interpreted as the percentage salary difference due to education.

In the model (5:7) the effect due to education and the effect due to job are additive, i.e. the effect due to education does not depend on the job and vice versa. The model is called an additive model. This is not always a realistic assumption. The ratio between the expected salaries of two educational qualifications may depend on the job or, which is the same thing, the ratio between the expected salaries of two jobs may depend on the education selected for the comparison. If this is the case there is interaction between the two factors education and job. To investigate interactions, effects due to each combination of education and job, γ_{ij}, can be added to the model. The expression (5:5) is then reformulated as

$$\ell_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}\,; \quad i=1,\ldots,p. \quad j=1,\ldots,q \qquad (5:11)$$

and (5:7) as

$$\ell_{ijk} = \mu + \sum_{i=1}^{p} \alpha_i x_i + \sum_{j=1}^{q} \beta_j y_j + \Sigma\gamma_{ij}(x_i y_j) + \epsilon_{ijk}\,; \qquad (5:12)$$

$(x_i y_j)$ can be considered as a member of a new set of dummy variables the value of which is unity when an observation belongs to education i *and* job j and is otherwise zero. The parameters in (5:11) and (5:12) are not unique. This can be seen for instance in the following way. Suppose there are pq unique expected values $\ln\phi_{ij}$.

$$\ln\phi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}\,; \quad i=1,\ldots,p. \quad j=1,\ldots,q \qquad (5:13)$$

There are thus pq independent equations but pq+p+q+1 unknown parameters. It is thus impossible to solve (5:13) for these parameters and to obtain a unique solution unless p+q+1 independent constraints are added to the system of equations. If (5:9) and (5:10) are used as before another p+q−1 constraints are needed. The new constraints are built up in the following way

$$\sum_{i=1}^{p} W_{ij}\gamma_{ij} = 0; \quad j=1,\ldots,q \qquad (5:14)$$

$$\sum_{j=1}^{q} W_{ij}\gamma_{ij} = 0; \quad i=1,\ldots,p \qquad (5:15)$$

These constraints are not linearly independent, one constraint is redundant and can be dropped. It is immaterial which one is chosen. There will then remain p+q–1 independent constraints.

The so called »main effects» α_i (and β_j) no longer tell the whole story in a comparison between two educational qualifications (jobs). If interactions are important it is necessary to specify the job (education) for which the comparison is done and to include the interaction effect γ_{ij}. For instance, for the educational qualifications i and r, the percentage difference in job j due to education is (approximately) $100(\alpha_i + \gamma_{ij} - \alpha_r - \gamma_{rj})$ per cent according to (5:11) and (5:12). $100(\alpha_i - \alpha_r)$ can now only be interpreted as an *average* difference due to education. The interpretation of the parameters is treated more thoroughly in section 5.1.2.2.

5.1.2 *The choice of constraints*

5.1.2.1 Formal properties of the constraints

The model (5:12) can be written in matrix form

$$\ell = Z\tau + \epsilon; \qquad (5:16)$$

where

$$\ell = \{\ell_{ijk}\}_{n.. \times 1}$$

$$Z = \left\{1, x_{1k},\ldots,x_{pk}, y_{1k},\ldots,y_{qk}, (x_1y_1)_k,\ldots,(x_1y_q)_k,\ldots,(x_py_q)_k\right\}_{n.. \times c} \quad n.. > c$$

$$\tau' = \left\{\mu, \alpha_1,\ldots,\alpha_p, \beta_1,\ldots,\beta_q, \gamma_{11},\ldots,\gamma_{1q}, \gamma_{pq}\right\}_{1 \times c}$$

$$\epsilon = \left\{\epsilon_{ijk}\right\}_{n.. \times 1}$$

With an expression borrowed from experimental statistics Z is the design matrix. The number of parameters (columns of Z) is c. With the notation previously introduced it is true that

$$c \leqslant 1 + p + q + pq; \qquad (5:17)$$

The inequality is valid when one or more interactions do not exist. Any model of the same kind as (5:12) can of course be written in matrix form in the same way as this model. The results obtained below are true for any model of the form (5:16) and are in particular not limited to a model with two factors.

The difference $\ell - \epsilon$ is now denoted g and (5:16) is written

$$Z\tau = g; \tag{5:18}$$

This implies that g is in the column space of Z. For the particular model in the previous section it was demonstrated that the columns of Z were linearly dependent. It will now be assumed that

$$\text{rank } Z = r < c; \tag{5:19}$$

where c is the number of columns of Z. Since g is in the column space of Z a solution exists (Pringle & Rayner [1971], theorem 1.6 p.9). It is well known (see for instance Pringle & Rayner [1971] p. 10) that the general solution of (5:18) is

$$\tau = Z^{-}g + (I - Z^{-}Z)k; \tag{5:20}$$

where k is an arbitrary c x 1 vector and Z^{-} a generalized inverse of Z.★ There is thus an infinite number of τ vectors (5:20) which satisfy (5:18). As has already been explained the non-uniqueness of τ is not satisfactory from the viewpoint of interpretation and presentation. We will therefore choose one of the possible values of τ which lends itself to a convenient interpretation. This will be done by imposing appropriate additional constraints on the parameter vector τ. Appropriate here means that for every g in the column space of Z the constraints and (5:18) give

a) one and only one of the solutions defined by (5:20)

and

b) that this solution can be interpreted in a meaningful and desired way.

Necessary and sufficient conditions for the first requirement will be considered in what remains of this section and the problem of interpretation will be treated in the following section.

The constraints chosen are of the following linear and homogeneous form★★

$$H\tau = 0; \tag{5:21}$$

★ In the terminology used by Pringle & Rayner [1971] Z^{-} is a g_1-inverse to Z, i.e. a non-unique matrix which satisfies $ZZ^{-}Z = Z$.

★★ The constraints do not necessarily have to be homogeneous, but this is customary in analyses of variance and the four specific constraints considered later have this property. No need has thus been felt for a generalization.

where the matrix H is of the order t x c. (5:18) and (5:21) together now form a new equation system

$$\begin{Bmatrix} Z \\ \hline H \end{Bmatrix} \tau = \begin{Bmatrix} g \\ \hline 0 \end{Bmatrix} \quad ; \tag{5:22}$$

It has been proved by several authors, for instance by Plackett [1950], Scheffé [1959] and Seber [1966] that necessary and sufficient conditions for the existence of a unique solution to the equation system (5:22) for every g in the column space of Z.

i) rank $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix} = c$;

ii) no linear combination of the rows of H except the null vector, is a linear combination of the rows of Z.

Yet another proof is given in appendix B. For matrices H which are chosen such that the two conditions are satisfied the unique solution of (5:22) is

$$\tau = (Z'Z + H'H)^{-1}Z'g; \tag{5:23}$$

For a proof see the references given above or appendix B.

For given matrices Z and H it may be difficult to investigate whether or not condition ii) is satisfied. (There is usually no problem in determining the rank of a matrix, and the investigation of condition i) does not therefore raise any major difficulties.) It is therefore useful to observe that Z and H are complementary.★ It then follows that the two conditions above are equivalent to the following two

i) rank $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix} = c$;

iii) rank Z + rank H = rank $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix}$;

For a proof see appendix B.

In order to test whether a given matrix H satisfies the conditions, one method is to determine the rank of the matrices Z, H and $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix}$ and to find out whether the sum of rank Z and rank H equals rank $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix}$ and the number of unknown parameters. With modern computer techniques such a test can be incorporated in a regression program (Cedheim & Klevmarken [1969]).

An alternative method is suggested by Plackett [1950]. The following results are stated without proof. As rank Z is less than c there exists D of order c x (c–r) and rank (c–r) such that ZD = 0. One possibility is to choose the matrix

★ According to the definition by Chipman [1964] two matrices A and B are complementary if a) A and B both have k columns and rank A + rank B = k, and b) the row spaces of A and B have only the origin in common.

$$D = \left[\begin{array}{c} Z_{rr}^{-1} \; Z_{r(c-r)} \\ \hline -I_{(c-r)} \end{array}\right] \; ; \qquad (5{:}24)$$

where Z_{rr}^{-1} is the inverse of a non-singular r x r submatrix Z_{rr} of Z. Since rank Z is r, such a matrix exists. $Z_{r(c-r)}$ is a matrix of the remaining elements in the same rows of Z used to form Z_{rr}. If H is a (c–r) x r matrix of rank (c–r) such that HD is non-singular, rank H is (c–r) and rank $\left\{\frac{Z}{H}\right\}$ is c which implies that rank H + rank Z = rank $\left\{\frac{Z}{H}\right\}$. (In addition to Plackett [1950] and [1960] a proof is also given by Seber [1966].) The method is thus first to find D such that ZD = 0 and then to test whether HD is non-singular. As this method presupposes that rank Z is known since otherwise a matrix H with exactly (c–r) rows cannot be chosen, and as the process of finding a D and then testing HD for singularity does not seem more convenient in computer work than the rank tests suggested above, this method does not offer any practical advantages.

Although we now have at least two methods of checking whether a given matrix H fulfils the necessary and sufficient conditions, the problem of choosing a matrix H for this test still remains. It is of course desirable that the H first chosen should fulfil the conditions. It is usually convenient to choose a matrix H with no more than c–r rows to avoid the sum of rank Z and rank H being greater than rank $\{Z' \vdots H'\}$. It is thus necessary to know r, i.e. rank Z which can usually be calculated fairly easily for the models used in this study. To see this we will investigate the rank of a sequence of models starting with the most simple one-factor model.

The regressor matrix of a *one-factor model* is represented in (5:25).

$$\begin{array}{c} \mu \;\; \alpha_1 \;\; \alpha_2 \; \ldots \; \alpha_p \\ \left\{\begin{bmatrix} 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 1 & 0 & 0 & \ldots & 1 \end{bmatrix}\right\} \end{array} ; \qquad (5{:}25)$$

In the following α_i is called the i-th level of factor α and it is assumed that there is one and only one observation on each factor level in the case of only one factor and on each factor combination in the case of more than one factor. This assumption is no limitation. If there is more than one observation on a factor combination the corresponding row in the regressor matrix is repeated as many times as there are observations; it does not change the rank of the matrix.

As (5:25) is of the order p x (p+1), rank (5:25) $\leqslant$ p. By inspection it is easy to find that the last p columns form an identity matrix. Thus rank (5:25) = p. The columns of the matrix are linearly dependent. The sum of the last p columns equals the first.

The regressor matrix for an additive *two-factor model* is obtained from (5:25) by a repetition of each row in this matrix once for each level of the new factor β. The α-regressors still add up to the first column and so do the β-regressors. As the columns in the new regressor matrix (5:26) are p+q+1 which is less than the number of rows, rank (5:26) $\leq$ p+q+1−2. To assure that there is equality

$$\begin{matrix} \mu & \alpha_1 & \alpha_2 & \dots & \alpha_p & \beta_1 & \beta_2 & \dots & \beta_q \end{matrix}$$
$$\begin{Bmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{Bmatrix} \qquad (5{:}26)$$

we first note that the rank of the matrix formed by the first p+1 columns is p (this matrix is the same as (5:25), except that each row is repeated q times which does not change its rank), and secondly that no linear combination of (q−1) of the β-regressors is also a linear combination of the μ- and α-regressors except the zero vector. Suppose for instance that the last β-regressor is dropped. A linear combination of the remaining (q−1) columns must always have a zero element in the q-th, 2q-th,– – –, and pq-th position and as it is impossible to form a linear combination of the first p+1 columns with this property, rank (5:26) = = p+q−1.

The same arguments can be repeated for any number of factors which gives the following *general* rule for an *additive model.*

$$\text{rank (regressor matrix)} = \sum_{i=1}^{f} (p_i - 1) + 1; \qquad (5{:}27)$$

where p_i is the number of levels of factor i and f the number of factors.

Consider now a *two-factor model with interactions* [(5:28)]·

$$
\begin{array}{c}
\begin{array}{ccccccccccccccccccccc}
\mu & \alpha_1 & \alpha_2 & \cdots & \alpha_p & \beta_1 & \beta_2 & \cdots & \beta_q & (\alpha\beta)_{11} & (\alpha\beta)_{12} & \cdots & (\alpha\beta)_{1q} & (\alpha\beta)_{21} & (\alpha\beta)_{22} & \cdots & (\alpha\beta)_{2q} & \cdots & (\alpha\beta)_{p1} & (\alpha\beta)_{p2} & \cdots & (\alpha\beta)_{pq}
\end{array}\\
\left\{\begin{array}{cccccccccccccccccccccc}
1 & 1 & & & & 1 & & \cdots & & 1 & & & & & & & & & & & & \\
1 & 1 & & & & & 1 & & & & 1 & & & & & & & & & & & \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
1 & 1 & & & & & & & 1 & & & & 1 & & & & & & & & & \\
1 & & 1 & & & 1 & & & & & & & & 1 & & & & & & & & \\
1 & & 1 & & & & 1 & & & & & & & & 1 & & & & & & & \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
1 & & 1 & & & & & & 1 & & & & & & & & 1 & & & & & \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
1 & & & & 1 & 1 & & & & & & & & & & & & & 1 & & & \\
1 & & & & 1 & & 1 & & & & & & & & & & & & & 1 & & \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\
1 & & & & 1 & & & & 1 & & & & & & & & & & & & & 1
\end{array}\right\}
\end{array}
\qquad (5{:}28)
$$

To simplify the notation all zero elements have been omitted from the matrix. By assumption (5:28) has pq rows and 1+p+q+pq columns and thus rank (5:28) $\leqslant$ pq. As the last pq columns form an identity matrix there is equality and rank (5:28) = pq. There are thus p+q+1 independent linear relations among the columns. The same two relations from the additive model still hold, namely that the sum of the α-regressors and the sum of the β-regressors each equal the μ-regressor. Furthermore, for each i the sum of the $(\alpha\beta)_{ij}$ columns over j equals the α_i column and for each j the sum of the $(\alpha\beta)_{ij}$ columns over all i equals the β_j column. However, this does not give p+q independent relations, but p+q−1; because one can be obtained from the others. This is seen from the following vector relations.

$$\mu = \sum_{i=1}^{p} \alpha_i \; ; \tag{5:29}$$

$$\beta_q = \mu - \sum_{j=1}^{q-1} \beta_j = \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q-1} \beta_j \; ; \tag{5:30}$$

$$\alpha_i = \sum_{j=1}^{q} (\alpha\beta)_{ij}; \qquad i=1,\ldots,p \tag{5:31}$$

$$\beta_j = \sum_{i=1}^{p} (\alpha\beta)_{ij}; \qquad j=1,\ldots,(q-1) \tag{5:32}$$

After substitution of (5:29) and (5:32) into (5:30) and a change of the summation order we obtain

$$\beta_q = \sum_{j=1}^{q} \sum_{i=1}^{p} (\alpha\beta)_{ij} - \sum_{j=1}^{q} \sum_{i=1}^{p} (\alpha\beta)_{ij} = \sum_{i=1}^{p} (\alpha\beta)_{iq} \; ; \tag{5:33}$$

p+q+1 independent linear relations have now been demonstrated.

The vector $\{\ln\phi_{ij}\}_{pq\times 1}$ of a two-factor model with interactions as (5:13) thus belongs to a pq-dimensional space, while a model without interactions is restricted to a p+q−1 dimensional space. The restrictions imposed on the additive model are of the following kind. For instance for p=q=2, (5:13) gives

$$\ln\phi_{ij} = \mu + \alpha_i + \beta_j \; ; \qquad i=1,2. \; j=1,2 \tag{5:34}$$

which yields the restriction

$$\ln\phi_{11} - \ln\phi_{12} = \ln\phi_{21} - \ln\phi_{22} \; ; \tag{5:35}$$

In this particular case, $p=q=2$, there is only one restriction on $\{\ln\phi_{ij}\}$ which belongs to a 3-dimensional space. In the general additive two-factor model there are $(p-1)(q-1)$ independent restrictions

$$\ln\phi_{ij} - \ln\phi_{kj} = \ln\phi_{i1} - \ln\phi_{k1} \; ; \qquad (5{:}36)$$

imposed on $\{\ln\phi_{ij}\}$ by the model.

The introduction of a *third factor,* first only in additive form, changes the regressor matrix of the model. Each row in (5:28) is repeated once for every level of the new factor. The regressors belonging to this factor add up to the μ-vector in analogy to the first two factors, but except for this linear relation the third factor does not introduce new linear relations among the columns. To see this, suppose the third factor has r levels and that one of its regressors is dropped, for instance the k-th one. Then the k-th, (k+r)-th, (k+2r)-th,– – – rows in the matrix, formed by the remaining k–1 regressors, only have zero elements. As it is impossible to obtain a linear combination of the first $1+p+q+pq$ regressors with zero elements in these rows, there is no linear combination of these regressors which is also a linear combination of the last $r-1$ regressors, except the zero vector. The rank of the regressor matrix with three factors and one two-factor interaction is thus $pq+r-1$.

Expansion of the model by a new two-factor interaction, for instance between the first and the third factor, adds pr columns to the regressor matrix. Analogously to the two-factor model there are $p+r-1$ independent relations between the new pr regressors and the old $1+p+q+r+pq$ regressors. There are no more linear relations which can be seen from the following example with $p=q=2$ and $r=3$ [(5:37)].

Suppose the first 6 and the last 4 regressors are dropped, the remaining 8 regressors are then linearly independent and the rank of the matrix is 8. The rank of a three-factor model with two two-factor interactions is $pq+pr-p =$
$= (p-1)(q-1) + (p-1)(r-1) + (p-1) + (q-1) + (r-1) + 1$.

Although more complicated, the same kinds of argument can be applied to models involving more factors and interactions. Our findings may then be generalized to a model with f factors and $s \leqslant \binom{f}{2}$ two-factor interactions

$$\text{rank (regressor matrix)} = 1 + \sum_{i=1}^{f}(p_i-1) + \sum_{\{ij\}=1}^{s}(p_i-1)(p_j-1); \qquad (5{:}38)$$

As the number of parameters m is

$$m = 1 + \sum_{i=1}^{f} p_i + \sum_{\{ij\}=1}^{s} p_i p_j \; ; \qquad (5{:}39)$$

μ	α_1	α_2	β_1	β_2	λ_1	λ_2	λ_3	$(\alpha\beta)_{11}$	$(\alpha\beta)_{12}$	$(\alpha\beta)_{21}$	$(\alpha\beta)_{22}$	$(\alpha\lambda)_{11}$	$(\alpha\lambda)_{12}$	$(\alpha\lambda)_{13}$	$(\alpha\lambda)_{21}$	$(\alpha\lambda)_{22}$	$(\alpha\lambda)_{23}$
1	1		1		1			1				1					
1	1		1			1		1					1				
1	1		1				1	1						1			
1	1			1	1				1			1					
1	1			1		1			1				1				
1	1			1			1		1					1			
1		1	1		1					1					1		
1		1	1			1				1						1	
1		1	1				1			1							1
1		1		1	1						1				1		
1		1		1		1					1					1	
1		1		1			1				1						1

(5:37)

the number of constraints required, c, is

$$c = m - \text{rank} = f + \sum_{\{ij\}=1}^{s} (p_i + p_j - 1); \tag{5:40}$$

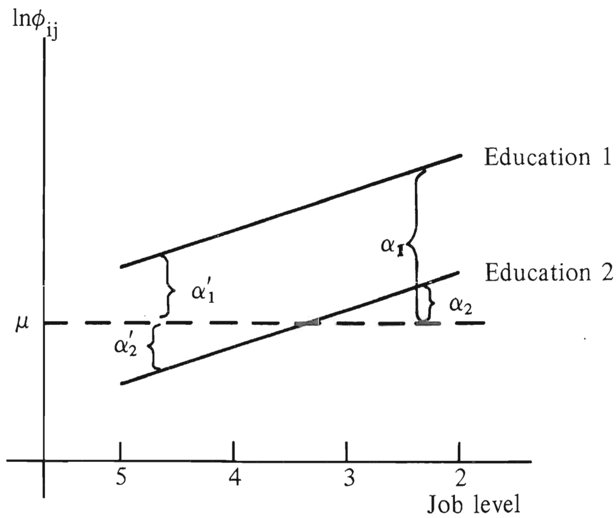
Even if the rules given are followed it may occur that the rankcheck indicates that rank $\{Z' \mid H'\}$ is less than the number of parameters and the sum of rank Z and rank H. This was a rare occurrence in our applications but when it happened it was usually possible to trace it back to a »numerical» singularity rather than a true singularity of $\{Z' \mid H'\}$. As will be shown in the next section the matrices H used in practical work are sometimes formed on the basis of the number of observations in each factor combination. Experience then indicates that numerical singularity may occur when the observations are distributed very unevenly.

It still remains to determine a particular set of constraints. The set of possible constraints is restricted to those which fulfil the two conditions i)and ii) on page 141 (or i) and iii) on page 141) and the choice cannot thus influence the rank of the model, only the interpretation of the parameters. In the next section four particular sets of constraints will be explored.

5.1.2.2 Four specific sets of constraints

In an additive model the effects due to one of the factors are independent of the other factors. This is illustrated in figure 5:1 for a model with two factors, education and job level. There are two kinds of education and four job levels.

Figure 5:1. *Main effects in an additive model*



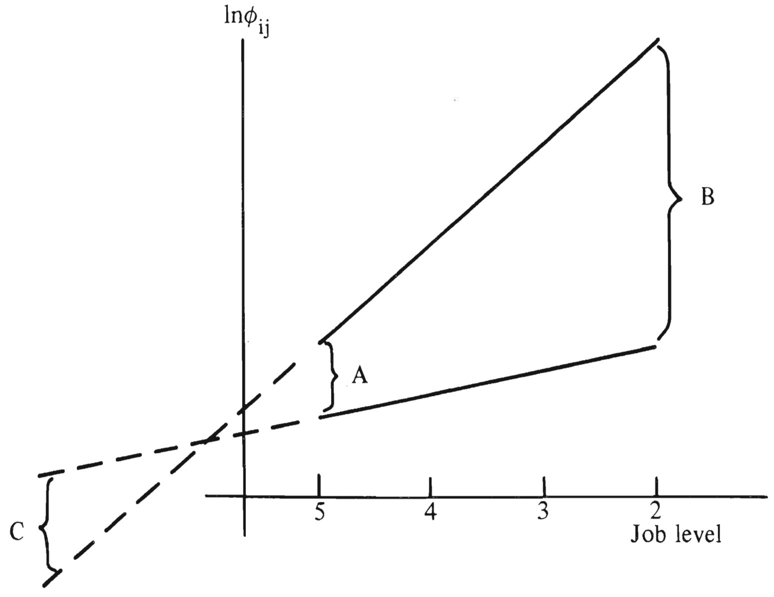
The upper solid curve shows the logarithmic salaries for employees with education 1 at various job levels. The lower curve shows the logarithmic salaries for employees with education 2. As the slope of the two curves in the diagram is positive the logarithmic salary increases by job level. The distance between the two curves is the same independently of job level as the model is additive, in this particular case they are two parallel lines. This distance then is a natural measure of the difference between the two educational effects.

The choice of constraints defines the overall average μ, (the line of short dashes). It also determines where along the job-level-axis the main effects are defined. If for instance, the constraints are chosen so that observations from high level jobs are given a heavy weight, the effect due to education 1 can be interpreted as a difference between the upper solid line and the line of short dashes in the upper job range, for example α_1, and similarly for the effect due to education 2, α_2. If the observations from low level jobs are weighted heavily the main effects may for instance be defined as α'_1 and α'_2. The latter then is negative. In both cases the difference between the two effects is the same.

The choice of constraints and the corresponding interpretation of the parameters may be of particular importance for models with interactions. Suppose for instance we have a two-factor model as before, but education and job level now interact as shown in figure 5:2. There is in this case no unique difference between the two kinds of education. What should properly be understood by a main effect therefore is unclear. If the difference between the two educational effects is measured at job level 5 the results is A but at level 2 the difference is larger, B. It is probably possible to choose constraints which define the main effects at job levels outside the present interval of levels and still satisfy the formal requirements. With constraints of this type the interpretation of the main effects would in most cases be rather strange. The difference C in figure 5:2 is one example, where even the sign in the comparison between the kinds of education is reversed. In cases with interactions it is usually satisfactory to define the difference between two main effects as some kind of an average of all possible differences inside the *present* range of categories, (job level 5–2 in this example).

In the following four sets of constraints will be suggested and their corresponding interpretation of the model parameters investigated and compared. The first set defines the main effects of education at a certain job level and vice versa, while the three remaining sets define the educational effects as averages over

Figure 5:2. *Main effects in a model with interactions*



job levels and vice versa. All of them are more or less common in practice. The first two have perhaps been applied most frequently in the analysis of non-experimental data (Suits [1957], Scheffé [1959], and Melichar [1965]), while the last is usually applied in the analysis of experimental data. By experience they have all been found to fulfil the conditions i) and ii) on page 141 provided that no combinations of categories are missing. The choice between them is primarily determined by the parameter interpretation desired. There are times when this criterion is not enough to single out only one set of constraints, but two or three alternatives may all give convenient interpretations. It is then relevant to investigate whether the estimates obtained by the different sets differ very much. If they give approximately the same result any of them will suffice. The interpretation of the parameters which follows from each of the four sets of constraints is investigated in a formal way in this section, while an empirical comparison is reserved for section 5.2.2.

For reasons of simplicity only the two-factor model is treated, but the results can be generalized. The same notation is used as before, see expressions (5:4) and (5:13).

The first set of constraints

$$\alpha_I = 0; \qquad (5:41)$$

$$\beta_J = 0; \qquad (5:42)$$

$$\gamma_{iJ} = 0; \quad i=1,\dots,p \qquad (5:43)$$

$$\gamma_{Ij} = 0; \quad j=1,\dots,q \qquad (5:44)$$

I is a particular education and J a particular job. Substitution of these constraints into (5:13) yields the following interpretation of the parameters.

$$\mu = \ln\phi_{IJ}; \qquad (5:45)$$

$$\alpha_i = \ln\phi_{iJ} - \ln\phi_{IJ}; \quad i=1,\dots,p \qquad (5:46)$$

$$\beta_j = \ln\phi_{Ij} - \ln\phi_{IJ}\ ; \quad j=1,\dots,q \qquad (5:47)$$

$$\gamma_{ij} = \ln\phi_{ij} - \ln\phi_{iJ} - \ln\phi_{Ij} + \ln\phi_{IJ}\ ; \quad i=1,\dots,p.\ j=1,\dots,q \qquad (5:48)$$

With these constraints μ is the expected (logarithmic) salary of education I and job J and all other parameters are interpreted as deviations from this expected value. The main effects α_i are determined by the (logarithmic) salary differences in job J only and the β_j's are determined by the (logarithmic) differences in education I only. This set of constraints is usually chosen when a particular education – job combination is a natural reference point, but when there is no such reference point and when the model contains interactions, this set may be somewhat peculiar in many applications. (When the model is additive all differences $\ln\phi_{iJ}-\ln\phi_{IJ}$; J=1,...,q, ($\ln\phi_{Ij}-\ln\phi_{IJ}$; I=1,...,p) are alike, see (5:36).) Sometimes it is more natural to define α_i as some kind of an average over all jobs and β_j as an average over all educational qualifications. The remaining three sets of constraints have this property.

The second set of constraints

$$\sum_{i=1}^{p} n_{i\bullet}\, \alpha_i = 0; \qquad (5:49)$$

$$\sum_{j=1}^{q} n_{\bullet j}\, \beta_j = 0; \qquad (5:50)$$

$$\sum_{i=1}^{p} n_{i\cdot}\gamma_{ij} = 0; \qquad j=1,\dots,q \tag{5:51}$$

$$\sum_{j=1}^{q} n_{\cdot j}\gamma_{ij} = 0; \qquad i=1,\dots,p \tag{5:52}$$

Application of these constraints to the model (5:13) yields

$$\sum_i\sum_j n_{i\cdot}n_{\cdot j}\ln\phi_{ij} = \sum_i\sum_j n_{i\cdot}n_{\cdot j}(\mu+\alpha_i+\beta_j+\gamma_{ij}) = \mu n_{\cdot\cdot}^2 + \sum_i n_{i\cdot}\alpha_i \sum_j n_{\cdot j} +$$

$$+ \sum_j n_{\cdot j}\beta_j \sum n_{i\cdot} + \sum_i n_{i\cdot}\sum_j n_{\cdot j}\gamma_{ij} = \mu n_{\cdot\cdot}^2 ; \tag{5:53}$$

$$\therefore\ \mu = \sum_i\sum_j \frac{n_{i\cdot}}{n_{\cdot\cdot}}\frac{n_{\cdot j}}{n_{\cdot\cdot}} \ln\phi_{ij}; \tag{5:54}$$

$$\sum_j n_{\cdot j}\ln\phi_{ij} = \sum_j n_{\cdot j}(\mu+\alpha_i+\beta_j+\gamma_{ij}) = \mu n_{\cdot\cdot} + \alpha_i n_{\cdot\cdot} ; \tag{5:55}$$

$$\therefore\ \alpha_i = \sum_j \frac{n_{\cdot j}}{n_{\cdot\cdot}} \ln\phi_{ij} - \mu; \qquad i=1,\dots,p \tag{5:56}$$

The expression β_j is obtained in the same way,

$$\therefore\ \beta_j = \sum_i \frac{n_{i\cdot}}{n_{\cdot\cdot}} \ln\phi_{ij} - \mu; \qquad j=1,\dots,q \tag{5:57}$$

and

$$\gamma_{ij} = \ln\phi_{ij} - \sum_j \frac{n_{\cdot j}}{n_{\cdot\cdot}}\ln\phi_{ij} - \sum_i \frac{n_{i\cdot}}{n_{\cdot\cdot}}\ln\phi_{ij} + \sum_i\sum_j \frac{n_{i\cdot}}{n_{\cdot\cdot}}\frac{n_{\cdot j}}{n_{\cdot\cdot}}\ln\phi_{ij}; \tag{5:58}$$

μ is now interpreted as a weighted average of expected (logarithmic) salaries where the weights are formed by the marginal sums. α_i is the average (logarithmic) salary difference of education i from the overall average salary. The average is again a weighted average in which the expected (logarithmic) salary in education i and job j is weighted proportional to the *total* number of employees in job j. A similar interpretation holds for the average (logarithmic) salary difference β_j due to job.

The third set of constraints

$$\sum_{i=1}^{p} n_{i.}\alpha_i = 0; \tag{5:59}$$

$$\sum_{j=1}^{q} n_{.j}\beta_j = 0; \tag{5:60}$$

$$\sum_{i=1}^{p} n_{ij}\gamma_{ij} = 0; \qquad j=1,\ldots,q \tag{5:61}$$

$$\sum_{j=1}^{q} n_{ij}\gamma_{ij} = 0; \qquad i=1,\ldots,p \tag{5:62}$$

The interpretation of the parameters μ, α_i, β_j and γ_{ij} associated with this new set of constraints is derived in the same way as before,

$$\sum_i\sum_j n_{ij}\ln\phi_{ij} = \sum_i\sum_j n_{ij}(\mu+\alpha_i+\beta_j+\gamma_{ij}) = \mu n_{..} + \sum_i \alpha_i n_{i.} + \sum_j \beta_j n_{.j} +$$

$$+ \sum_i\sum_j n_{ij}\gamma_{ij} = \mu n_{..}\,; \tag{5:63}$$

$$\therefore\ \mu = \sum_i\sum_j \frac{n_{ij}}{n_{..}} \ln\phi_{ij}; \tag{5:64}$$

$$\sum_j n_{ij}\ln\phi_{ij} = \sum_j n_{ij}(\mu+\alpha_i+\beta_j+\gamma_{ij}) = \mu n_{i.} + \alpha_i n_{i.} + \sum_j n_{ij}\beta_j; \tag{5:65}$$

$$\therefore\ \alpha_i = \sum_j \frac{n_{ij}}{n_{i.}}(\ln\phi_{ij}-\beta_j) - \mu; \qquad i=1,\ldots,p \tag{5:66}$$

Analogously

$$\beta_j = \sum_i \frac{n_{ij}}{n_{.j}}(\ln\phi_{ij}-\alpha_i) - \mu; \qquad i=1,\ldots,q \tag{5:67}$$

With the constraints (5:59) – (5:62) also, μ is a weighted average of all expected (logarithmic) salaries, but the weights are now proportional to the number of employees in each combination of education and job. The main effects are a weighted average of (logarithmic) differences from the overall mean as in the previous case, but before the averages giving the educational effects are formed, each expected (logarithmic) salary is adjusted for the job effect, and before the averages giving the job effects are formed, each expected (logarithmic) salary is adjusted for the educational effect.

The fourth set of constraints

$$\sum_{i=1}^{p} \alpha_i = 0; \tag{5:68}$$

$$\sum_{j=1}^{q} \beta_j = 0; \tag{5:69}$$

$$\sum_{i=1}^{p} \gamma_{ij} = 0; \qquad j=1,\ldots,q \tag{5:70}$$

$$\sum_{j=1}^{q} \gamma_{ij} = 0; \qquad i=1,\ldots,p \tag{5:71}$$

(5:13) and (5:68) – (5:71) give

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \ln\phi_{ij} = \sum_{i=1}^{p} \sum_{j=1}^{q} (\mu+\alpha_i+\beta_j+\gamma_{ij}) = \mu pq; \tag{5:72}$$

$$\because \ \mu = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} \ln\phi_{ij}; \tag{5:73}$$

$$\sum_{j=1}^{q} \ln\phi_{ij} = \sum_{j=1}^{q} (\mu+\alpha_i+\beta_j+\gamma_{ij}) = \mu q + \alpha_i q; \tag{5:74}$$

$$\because \ \alpha_i = \frac{1}{q} \sum_{j=1}^{q} \ln\phi_{ij} - \mu; \qquad i=1,\ldots,p \tag{5:75}$$

and

$$\beta_j = \frac{1}{p} \sum_{i=1}^{p} \ln\phi_{ij} - \mu; \qquad j=1,\ldots,q \tag{5:76}$$

This fourth set of constraints gives an interpretation similar to the second set. The only difference is that the averaging is done with equal weights.

The following numerical example is chosen to illustrate the properties of the four sets of constraints. Suppose there are three educational qualifications and three jobs. The expected logarithmic salaries are given in table 5:1. To demonstrate the importance of the distribution of employees by education and job, alternative distributions are given in table 5:2. The overall mean and the educational main effects are calculated for each set of constraints and with each alternative in table 5:2. Table 5:3 contains the results from these calculations.

Table 5:1. *Expected logarithmic salaries*

Education	Job 1	2	3
1	6.0	7.5	8.0
2	6.0	7.1	7.2
3	8.0	7.5	7.0

Table 5:2. *Number of employees by education and job*

Education	Alternative A				Alternative B			
	Job				Job			
	1	2	3	Σ	1	2	3	Σ
1	10	20	30	60	10	20	30	60
2	8	7	15	30	5	7	18	30
3	2	3	5	10	5	3	2	10
Σ	20	30	50	100	20	30	50	100

Table 5:3. *Educational main effects*

Set of con-straints	Results obtained with							
	Alternative A				Alternative B			
	μ	α_1	α_2	α_3	μ	α_1	α_2	α_3
1★	6.0	0.0	0.0	2.0	6.0	0.0	0.0	2.0
2	7.464	0.186	−0.334	−0.114	7.464	0.186	−0.334	−0.114
3	7.472	0.188	−0.334	−0.123	7.508	0.153	−0.405	0.296
4	7.367	0.133	−0.267	0.133	7.367	0.133	−0.267	0.133

★ Reference point is education 1 and job 1, i.e. I=J=1.

With the first set of constraints the salary of the first education and the first job combined is chosen as the reference point, therefore μ equals the salary of this combination. The effects due to education are then determined as the differences between the first and the other educational groups in the first job. If some other reference point is chosen the main effects will in general be different. The first education and the third job would for instance give the effects due to education: 0, –0.8, –1.0. This may seem a little arbitrary. The three other constraints do not have this property, but two of them, the second and the third, depend on the number of employees, while the first and fourth do not.

Given that the numbers in table 5:1 are natural logarithms the parameters in table 5:3 can roughly be interpreted as percentage differences. For instance with set No. 2 and alternative A the average salary in education 1 is approximately 18.6 % above the overall average salary. In education 2 it is 33.4 % below and in education 3, 11.4 % below.★

The two alternatives in table 5:2 are made up in such a way that they have the same marginal distributions. The difference between A and B is only that observations are shifted from job 1 to job 3 in education 2 and from 3 to 1 in education 3. As the second set of constraints only depends on the marginal distributions the application of these constraints yields the same result with alternative A as with B. As set No. 4 does not depend at all on the distribution of employees, the parameters associated with this set are also repeated twice in table 5:3. It is, however, instructive to compare the parameters associated with sets Nos.2 and 4. The effect due to education 3 is positive with set No. 4 but negative with set No. 2. The explanation of the negative sign is of course that 50 % of the employees have job 3, where education 3 is given a relatively low reward. Although only 5 % (2 %) of the employees have this combination, its low salary is weighted heavily with the second set of constraints. Set No. 3 behaves differently in alternative B but not in A. In A the conditional distributions are very close to the marginal distributions and sets Nos.2 and 3 thus give almost the same result, but in alternative B the distribution conditional on education 3 is positively instead of negatively skew and the weight applied to the salary in job 3 with education 3 is very much reduced which gives a positive effect due to education 3.

The choice of constraints can thus be very essential to the interpretation of the model, but perhaps not as important as this numerical example may indicate (see below). In this study we use populations of employees and the distribution of employees determining the constraints is thus that of the pop-

---

★ The correct percentages are 20.4, 39.7 and 12.0, respectively.

ulation. In other applications observations may have been obtained by a sample survey. In that case it is less obvious that the constraints can be based on the distribution of employees *in the sample*, in particular if this empirical distribution is the result of the sampling experiment.

In some applications the comparison inherent in the constraints chosen is not the only matter of interest. Sometimes it is also interesting to investigate contrasts. The average salary difference between two educational qualifications or two jobs or the salary range due to education may for instance be the particular interest. In general contrasts are *not indifferent* to the choice of constraints. For instance, from (5:56) we obtain

$$\alpha_i^{(2)} - \alpha_I^{(2)} = \sum_{j=1}^{q} \frac{n_{\cdot j}}{n_{\cdot\cdot}} (\ln\phi_{ij} - \ln\phi_{Ij}); \tag{5:77}$$

and from (5:75)

$$\alpha_i^{(4)} - \alpha_I^{(4)} = \frac{1}{q} \sum_{j=1}^{q} (\ln\phi_{ij} - \ln\phi_{Ij}); \tag{5:78}$$

(5:77) and (5:78) do not in general coincide, but in one important special case they do, namely when the model is additive. This follows from (5:36).

From (5:66) we obtain

$$\alpha_i^{(3)} - \alpha_I^{(3)} = \sum_{j=1}^{q} \left(\frac{n_{ij}}{n_{i\cdot}} \ln\phi_{ij} - \frac{n_{Ij}}{n_{I\cdot}} \ln\phi_{Ij}\right) - \sum_{j=1}^{q} \beta_j \left(\frac{n_{ij}}{n_{i\cdot}} - \frac{n_{Ij}}{n_{I\cdot}}\right); \tag{5:79}$$

When the proportionality relation

$$\frac{n_{ij}}{n_{i\cdot}} = \frac{n_{Ij}}{n_{I\cdot}} \; ; \qquad j=1,\ldots,q \tag{5:80}$$

holds, (5:79) equals (5:77), but in general it does not.

If there is interaction between two factors, for instance education and job, the salary differences between two educational qualifications depend on the job.

$$(\alpha_I + \gamma_{Ij}) - (\alpha_i + \gamma_{ij}) = \ln\phi_{Ij} - \ln\phi_{ij}; \tag{5:81}$$

but the comparison is independent of the weights.

5.1.2.3 On the treatment of »empty cells»

Although not explicitly stated it has been assumed that all two-by-two factor combinations exist, i.e. it is possible to observe a salary for any combination of two factors. If this is not so and there are »empty cells» the number of parameters is reduced. Formally this can be accomplished by adding one or more constraints to the model. It has been demonstrated previously that the complete two-factor model (5:13) has a regressor matrix of rank pq. This matrix then contains pq independent rows. Each of them corresponds to one combination of the two factors and is repeated as many times as there are observations on this combination. Suppose now that combination (I, J) does not exist.★ It is then impossible to determine the interaction γ_{IJ}, and the rank of the regressor matrix is reduced by one as one of the previously pq independent rows is now missing. Formally, to obtain full rank, the following constraint may be added to the model

$$\gamma_{IJ} = 0; \qquad (5:82)$$

If there is more than one empty cell more constraints of the same type are added. Depending on where the empty cells are located these additional constraints may imply that even more parameters are zero. Assume for instance a complete two-factor model where each factor gives a maximum of three effects. Suppose now that the two combinations (1,2) and (3,2) are missing. Under any of the last three sets of constraints the main effect β_2 is in principle determined as the deviation between the overall average and the average of the expected salaries in the three combinations (1,2), (2,2) and (3,2), but with (1,2) and (3,2) missing it is only determined by the overall average and (2,2). Obviously no unique interaction γ_{22} can be determined in addition to β_2. With the constraints (5:49) – (5:52), (5:59) – (5:62) or (5:68) – (5:71) and $\gamma_{12} = \gamma_{32} = 0$ no additional constraint is needed because they imply that $\gamma_{22} = 0$. However, if it is instead combinations (1,1) and (3,3) which are the empty cells, adding $\gamma_{11} = \gamma_{33} = 0$ to the model does not imply that any other parameter is zero. The analysis of empty cell cases can be extended to models involving more factors and more effects, but there does not seem to exist any simple formula for the determination of what parameters become zero when there are empty cells.

Finally, it should be made clear that no specification error is committed when one or more parameters are dropped because certain factor combinations cannot exist, but in empirical work observational empty cells sometimes occur, i.e. no

★ Even if the constraints are not based on the number of observations, it may be useful to count the number of observations in each factor combination in order to discover if any combination lacks observations.

observations are actually obtained, but in principle it is possible to observe the cell. In this case it is not possible to estimate a parameter associated only with this cell, but the involuntary omission may bias the estimates of the other parameters.

5.1.3 *Estimation*

5.1.3.1 Least squares estimators

The two previous sections of this chapter have been used to analyse the properties of the theoretical models and, with some exceptions, estimation has not been touched upon. As the models used belong to the class of general linear models it is natural to apply least squares theory to estimate the model. In section 5.1.2 it was stated that if the constraints are chosen as to fulfil conditions i) and ii) on page 141 there is a unique parameter vector τ which satisfies (5:18) and (5:21). It must also be shown that there exists a unique least squares estimate $\hat{\tau}$ satisfying (5:21), i.e. that the constraints imposed on the parameters can also be imposed on the estimates. However, this immediately follows from the same statement on page 141 (proposition B:1 in Appendix B) if g is made equal to the orthogonal projection of ℓ on the linear space spanned by the columns of Z, i.e. the least squares estimate of ℓ. (See also Scheffé [1959] Corollary 2 to theorem 3.)

Suppose $\hat{\tau}$ is the unique least squares estimate. Thus it is one of the solutions to the normal equations and satisfies

$$Z'\ell = Z'Z\hat{\tau}; \tag{5:83}$$

and as it satisfies (5:21) it is also true that

$$H'H\hat{\tau} = 0; \tag{5:84}$$

Adding (5:83) and (5:84) we note that

$$Z'\ell = \left\{Z' \mid H\right\}\left\{\frac{Z}{H}\right\}\hat{\tau}\ ; \tag{5:85}$$

By assumption $\{Z' \mid H'\}$ has full rank. From a well known theorem of matrix algebra it then follows that $\{Z' \mid H'\}\left\{\frac{Z}{H}\right\}$ also has full rank and that its inverse exists. The least squares estimator is then obtained from (5:85).

$$\hat{\tau} = (Z'Z + H'H)^{-1}Z'\ell; \tag{5:86}$$

If the model (5:16) and the homogeneous condition (5:21) are substituted into (5:86) it is easily shown that $\hat{\tau}$ is a conditional unbiased estimate. The condition is the set of constraints.

If the moment matrix $E(\epsilon\epsilon')$ takes the simple form $\sigma^2 I$, where σ^2 is a constan it follows from the Gauss-Markov theorem (see for instance Malinvaud [1966], p. 151) that $\hat{\tau}$ is the best linear conditional unbiased estimate. The variance-covariance matrix of $\hat{\tau}$ is as follows (for proof see Plackett [1960], p. 43 or Sebe [1966], p. 18).

$$E(\hat{\tau}-\tau)(\hat{\tau}-\tau)' = \sigma^2(Z'Z + H'H)^{-1} Z'Z(Z'Z + H'H)^{-1} \; ; \qquad (5:87)$$

Without specifying the specific features of the model it is of course difficult to discuss the realism of this simple moment matrix. In the previous chapter it was shown that the individual salary variability increased by age. In the followin other variables will be used in addition to age which will tend to create more homogeneous subgroups of employees. The heteroscedasticity previously observe would now at least partly be explained by these new variables. We may also obtain some comfort from the fact that ordinary least squares estimates do not become biased, only inefficient, by heteroscedasticity.

The estimation procedure can technically be carried out in two different ways both of which however give the same result. One method is to solve the s independe constraints (5:21) for s parameters and to substitute them out of the model. The result is a reduced linear model with c–s new independent regressors which are linear forms of the old regressors (Melichar [1965]). The remaining parameters can be estimated by usual least squares regression techniques. By substitution back into the constraints estimates are obtained for the s redundant parameters. When a standard regression program is used for estimation, this method does not automatically provide standard errors of the parameter estimates. Another drawback of the method is that it may be difficult to formulate and program substitution routines general enough to cope with models with interactions and empty cells. Another method is then to add the constraints to the model as new observations and apply the ordinary regression – least squares technique. When the constraints (5:21) are added to the model (5:16) we obtain

$$\begin{Bmatrix} \ell \\ 0 \end{Bmatrix} = \begin{Bmatrix} Z \\ H \end{Bmatrix} \tau + \begin{Bmatrix} \epsilon \\ 0 \end{Bmatrix} \; ; \qquad (5:88)$$

and it follows immediately that the ordinary least squares estimator applied to (5:88) is nothing but the estimator (5:86).

5.1.3.2 Transformations from one set of estimates to another

It was previously mentioned that the desired standardization of parameters and estimates may change from one application to another. This raises the question whether it is possible to transform the estimates (and the parameters) from one standardization to another in order to avoid a complete re-estimation. As a matter of fact there is a transformation matrix T such that

$$\hat{\tau}_2 = T\hat{\tau}_1 \; ; \tag{5:89}$$

where $\hat{\tau}_1$ and $\hat{\tau}_2$ are two solutions of the normal equations (5:83)

$$\hat{\tau}_1 = (Z'Z + H_1'H_1)^{-1}Z'\ell; \tag{5:90}$$

$$\hat{\tau}_2 = (Z'Z + H_2'H_2)^{-1}Z'\ell; \tag{5:91}$$

and H_1 and H_2 are two matrices which both fulfil the conditions i) and ii) on page 141. The transformation matrix is obviously

$$T = (Z'Z + H_2'H_2)^{-1}(Z'Z + H_1'H_1); \tag{5:92}$$

Although T gives the transformation from one set of estimates to another, it does not in general offer any great computational advantage as the inverse of $(Z'Z + H_2'H_2)$ has to be calculated.

There may be an alternative approach based on the formulation of the general solution of the normal equations suggested in Lundquist [1970]★

$$\tau = \begin{Bmatrix} (Z_{nr}'Z_{nr})^{-1}Z_{nr}'\ell - (Z_{nr}'Z_{nr})^{-1}Z_{nr}'Z_{n(c-r)}\,k \\ k \end{Bmatrix}; \tag{5:93}$$

where Z_{nr} is an n x r matrix of r linearly independent column vectors of Z, $Z_{n(c-r)}$ is an n x (c–r) matrix of the remaining columns and k an arbitrary vector with (c–r) elements. The device is to choose k in such a way that the desired standardization is obtained. The only problem is that one has to know what k to choose in order to obtain a particular standardization. Although it is an interesting problem it is not necessary to pursue it in this study, and it is left for a future analysis.

★ I am indebted to S. Lundquist who pointed out this possibility to me.

5.1.3.3 Estimation with many regressors

To make the exposition easy the two-factor model has frequently been used as an illustration, but in practical work models with several factors are used. The number of dummy variables then easily increases beyond the limits set by ordinary standard programs and by core storages. This is in particular true for models with interactions. How serious this problem is depends of course on the computer equipment available. With modern computers it is not primarily a technical or a systems problem, but more an economic one.[*] At the time when most of the empirical work reported in section 5.2 was done, it was serious enough. However, the rapid development of computer techniques has probably made this problem of less immediate interest today, but it is not completely without interest. Everybody has not access to a large-scale computer and even so the desire to build large models may make the computations unmanageable. Three different approaches to the problem will therefore be discussed in the following, not only for historical reasons. None of these alternatives really solves the problem, but offers second best solutions.

a) The set of observations is partitioned into two or more subsets and the model is estimated for each subset. The partitioning is done by one of the factors. For instance, to reduce the number of dummy variables for the job factor, employees engaged in production work may be analysed separately from those who hold positions in marketing divisions. Sometimes there is a close correlation between two factors which can be utilized to reduce the number of dummy variables even further. Some educational qualifications may for instanc predestinate for certain jobs and a partitioning by education will then at the same time imply a partitioning by job.

Suppose now that the model (5:88) can be partitioned in the following way

$$\begin{bmatrix} \ell_1 \\ \hdashline \ell_2 \\ \hdashline 0 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & 0 \\ Z_{21} & 0 & Z_{23} \\ H_1 & H_2 & H_3 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ 0 \end{bmatrix} ; \qquad (5{:}94)$$

As this model is assumed to have too many dummy variables to fit into the available computer the following two support models are estimated

---

[*] An interesting approach is suggested in Andrews, Morgan and Sonquist [1969], where the normal equations are solved by a simple iterative procedure. The data program (MCA) developed also contains some new measures of the »importance» of a factor which could be used as alternatives to the »factor ranges» used in this study (section 5.2). According to th authors: »MCA's major advantage is in taking the data the way they usually come, and prin ing out the results one is most likely to want to present and in a convenient way.» (p.128)

$$\begin{Bmatrix} \ell_1 \\ \hdashline 0 \end{Bmatrix} = \begin{Bmatrix} Z_{11} & \vdots & Z_{12} \\ \hdashline H_{11} & \vdots & H_{12} \end{Bmatrix} \begin{Bmatrix} \tau_{11} \\ \hdashline \tau_{12} \end{Bmatrix} + \begin{Bmatrix} \epsilon_1 \\ \hdashline 0 \end{Bmatrix} ; \qquad (5{:}95)$$

$$\begin{Bmatrix} \ell_2 \\ \hdashline 0 \end{Bmatrix} = \begin{Bmatrix} Z_{21} & \vdots & Z_{23} \\ \hdashline H_{21} & \vdots & H_{23} \end{Bmatrix} \begin{Bmatrix} \tau_{21} \\ \hdashline \tau_{23} \end{Bmatrix} + \begin{Bmatrix} \epsilon_2 \\ \hdashline 0 \end{Bmatrix} ; \qquad (5{:}96)$$

In appendix C it is shown that if $H_{11} = H_{21} = H_1$; $H_{12} = H_2$ and $H_{23} = H_3$ the least squares estimate (LSE) of an element of τ_1 is a weighted sum of the elements of the LSE of τ_{11} and τ_{21},

$$\hat{\tau}_{1J} = \omega_{11J}\hat{\tau}_{11J} + \omega_{21J}\hat{\tau}_{21J} + \sum_{\substack{k \\ k \neq J}} (\omega_{11k}\hat{\tau}_{11k} + \omega_{21k}\hat{\tau}_{21k}); \quad J=1,\ldots,k_1 \qquad (5{:}97)$$

and that the LSE of τ_2^1 (τ_3) equals the LSE of τ_{12} (τ_{23}) plus a weighted sum of the LSE of τ_{11} and τ_{21},

$$\hat{\tau}_{2J} = \hat{\tau}_{12J} + \sum_{k=1}^{k_1} (\omega^*_{11k}\hat{\tau}_{11k} + \omega^*_{21k}\hat{\tau}_{21k}); \quad J=1,\ldots,k_2 \qquad (5{:}98)$$

$$\hat{\tau}_{3J} = \hat{\tau}_{23J} + \sum_{k=1}^{k_1} (\omega^{**}_{11k}\hat{\tau}_{11k} + \omega^{**}_{21k}\hat{\tau}_{21k}); \quad J=1,\ldots,k_3 \qquad (5{:}99)$$

where the weights ω, ω^* and ω^{**} have the properties

$$\omega_{11J} + \omega_{21J} = 1; \quad J=1,\ldots,k_1 \qquad (5{:}100)$$

$$\omega_{11k} + \omega_{21k} = 0; \quad k \neq J \qquad (5{:}101)$$

$$\omega^*_{11k} + \omega^*_{21k} = 0; \quad k=1,\ldots,k_1 \qquad (5{:}102)$$

$$\omega^{**}_{11k} + \omega^{**}_{21k} = 0; \quad k=1,\ldots,k_1 \qquad (5{:}103)$$

(For a detailed explanation of the notation see appendix C.)

The same result was obtained in Klevmarken [1968b] for the special case $H_1 = H_{11} = H_{21} = 0$, an example of which is when $\begin{Bmatrix} Z_{11} \\ \hdashline Z_{21} \end{Bmatrix}$ is a vector with all elements equal to unity and τ_1 is the common intercept.

From (5:97) – (5:103) it follows that if

$$\hat{\tau}_{11} = \hat{\tau}_{21} \; ; \tag{5:104}$$

then $\hat{\tau}_1 = \hat{\tau}_{11} = \hat{\tau}_{21}$; (5:105)

and $\hat{\tau}_2 = \hat{\tau}_{12}$; (5:106)

and $\hat{\tau}_3 = \hat{\tau}_{23}$; (5:107)

The assumption that the estimates of the parameters common to the two support models are equal is rarely fulfilled in practice. The result (5:105) – – (5:107) may, however, indicate that in the case of small differences between $\hat{\tau}_{11}$ and $\hat{\tau}_{21}$, $\hat{\tau}_{12}$ and $\hat{\tau}_{23}$ are rather good estimates of τ_2 and τ_3 respectively. This interpretation should, however, be done with some caution, because although the weights add up to zero they do not individually have to be small, and even a small divergence from equality may give a large contribution to the estimate of τ_i.

The estimates $\hat{\tau}_1$, $\hat{\tau}_2$ and $\hat{\tau}_3$ are least squares estimates of τ_1, τ_2 and τ_3 respectively and under the usual assumptions about the stochastic error term they are best linear unbiased estimates and so are then the estimates obtained by (5:97), (5:98) and (5:99).

$\hat{\tau}_{11}$ and $\hat{\tau}_{21}$ are in general not even unbiased estimates of τ_1 nor are $\hat{\tau}_{12}$ and $\hat{\tau}_{23}$ unbiased estimates of τ_2 and τ_3. Suppose the constraints satisfy the conditions i) and ii) on page 141. The estimates $\hat{\tau}_{11}$ and $\hat{\tau}_{12}$ can then be transformed to the following form by the use of (5:94) and the assumption $H_{11} = H_{21} = H_1$; $H_{12} = H_2$ and $H_{23} = H_3$.

$$\begin{bmatrix} \hat{\tau}_{11} \\ \hat{\tau}_{12} \end{bmatrix} = \left[\begin{pmatrix} Z_{11} & Z_{12} \\ H_{11} & H_{12} \end{pmatrix}' \begin{pmatrix} Z_{11} & Z_{12} \\ H_{11} & H_{12} \end{pmatrix} \right]^{-1} \begin{pmatrix} Z_{11} & Z_{12} \\ H_{11} & H_{12} \end{pmatrix}' \left[\begin{pmatrix} Z_{11} & Z_{12} \\ H_1 & H_2 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} 0 \\ H_3 \end{pmatrix} \tau_3 + \begin{pmatrix} \epsilon_1 \\ 0 \end{pmatrix} \right]$$

$$= \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} + \left\{ \begin{pmatrix} Z_{11} & Z_{12} \\ H_1 & H_2 \end{pmatrix}' \begin{pmatrix} Z_{11} & Z_{12} \\ H_1 & H_2 \end{pmatrix} \right\}^{-1} \begin{pmatrix} Z_{11} & Z_{12} \\ H_1 & H_2 \end{pmatrix}' \left\{ \begin{pmatrix} 0 \\ H_3 \end{pmatrix} \tau_3 + \begin{pmatrix} \epsilon_1 \\ 0 \end{pmatrix} \right\} ; \tag{5:108}$$

Taking expectation of both sides reveals the bias component

$$\begin{Bmatrix} E(\hat{\tau}_{11} - \tau_1) \\ E(\hat{\tau}_{12} - \tau_2) \end{Bmatrix} = \left\{ \begin{pmatrix} Z_{11} & Z_{12} \\ H_1 & H_2 \end{pmatrix}' \begin{pmatrix} Z_{11} & Z_{12} \\ H_1 & H_2 \end{pmatrix} \right\}^{-1} \begin{Bmatrix} H_1' H_3 \\ H_1' H_3 \end{Bmatrix} \tau_3 \; ; \tag{5:109}$$

The bias vanishes in two special cases, namely when $H_3\tau_3 = 0$ or when the columns of H_3 are orthogonal to the columns of H_1 and H_2. Similar results can easily be obtained for $\hat{\tau}_{21}$ and $\hat{\tau}_{23}$. They are unbiased if $H_2\tau_2 = 0$ or if H_2 is orthogonal to H_1 and H_3. These conditions may or may not be satisfied. For instance, the first of the four sets of constraints considered previously satisfies both conditions, while the last three do not. In an additive model, partitioned such that the elements of τ_2 and τ_3 belong to the same factor and no element of τ_1 belongs to it, the first set satisfies $H_1'H_2 = 0$, $H_1'H_3 = 0$ and $H_2'H_3 = 0$, while the last three do not satisfy $H_2'H_3 = 0$.

b) A second alternative is to omit one factor (or interaction), estimate the truncated model, retrieve the factor and omit another, estimate the new truncated model, and so forth until all factors have been estimated at least once. This alternative differs from the previous one by the omission of *all* regressors of a factor (interaction). The effect of dropping one or more variables from a regression is disentagled in ordinary regression analysis (see for instance Malinvaud [1966]). The estimates obtained from the truncated model are the same as those which would have been obtained from the complete model, except for the common intercept, if the omitted and non-omitted regressors are uncorrelated.

In our case the model can be partitioned in the following way

$$\begin{Bmatrix} \ell \\ \hline 0 \\ \hline 0 \end{Bmatrix} = \begin{Bmatrix} i & Z_1 & Z_2 \\ 0 & H_1 & 0 \\ 0 & 0 & H_2 \end{Bmatrix} \begin{Bmatrix} \tau_0 \\ \hline \tau_1 \\ \hline \tau_2 \end{Bmatrix} + \begin{Bmatrix} \epsilon \\ \hline 0 \\ \hline 0 \end{Bmatrix} \; ; \qquad (5:110)$$

where i is a vector of unit elements, Z_1 an n x p matrix of all regressors belonging to one or more factors (interactions) and Z_2 an n x q matrix of all regressors belonging to one or more other factors (interactions). τ_0 is the common intercept and τ_1 and τ_2 are factor (interaction) parameters. After elimination of τ_0 the normal equations to (5:110) can be written as

$$\begin{Bmatrix} Z_1'(\ell - i\bar{y}) \\ \hline Z_2'(\ell - i\bar{y}) \end{Bmatrix} = \begin{Bmatrix} Z_1'(I-\frac{ii'}{n})Z_1 + H_1'H_1 & Z_1'(I-\frac{ii'}{n})Z_2 \\ Z_2'(I-\frac{ii'}{n})Z_1 & Z_2'(I-\frac{ii'}{n})Z_2 + H_2'H_2 \end{Bmatrix} \begin{Bmatrix} \tau_1 \\ \hline \tau_2 \end{Bmatrix} \; ; \qquad (5:111)$$

If H_1 and H_2 are such that the inverse of the matrices in the diagonal of the matrix to the right of the equality sign in (5:111) exists, it is easy to show that the estimate of τ_1 is the same whether or not Z_2 and H_2 are omitted, if

$$Z_1' (I - \frac{ii'}{n}) Z_2 = 0; \qquad (5:112)$$

i.e. if the columns of Z_1 are uncorrelated with the columns of Z_2. Suppose that the columns of Z_1 belong to only one factor and that the same is true for Z_2.* Assume also that subindex r stands for the r-th column of Z_1 and t for the t-th column of Z_2. As the elements of the matrix $Z_1' Z_2$ are the number of observations in each combination of the two factors, n_{rt}, the condition (5:12) can be rewritten as

$$n_{rt} = n \frac{n_{r\cdot}}{n} \frac{n_{\cdot t}}{n} \; ; \qquad r=1,\ldots,p. \quad t=1,\ldots,q \qquad (5:113)$$

This expression is of course nothing but the traditional formulation of the orthogonality condition in analyses of variance.

The conclusion drawn for practical work is obviously that it is only factors (interactions) with regressors that may be expected at least approximately to be uncorrelated with regressors of other factors and interactions which should be omitted. With reference to the empirical analyses in section 5.2 where effects due to age, job, cost of living area, industry and education are investigated, the possibility of omitting, for instance, the age factor is less likely than that of omitting the factors cost of living area and industry, because age is probably highly correlated with job level, while there is no immediate reason to expect a strong correlation between cost of living area, industry and other factors.

c) A third alternative is to use a priori information to constrain the parameters more than is formally necessary and thereby reduce the number of parameters. It is of course always desirable to use any a priori information to increase the precision of the estimates, but in this case when some measure is necessary to reduce the number of parameters one may be more willing than otherwise to use more uncertain information and qualified guesses.

In the empirical section below this method has been applied to reduce the number of interaction parameters. When for instance age interacts with another factor the effects due to adjacent age intervals are two by two assumed to be the same.

---

* For the second and third sets of constraints previously considered it is then true that

$i'Z_1 = H_1$ and $i'Z_2 = H_2$;

which gives the following alternative formulation of condition (5:112)

$$Z_1' Z_2 = \frac{H_1' H_2}{n} \; ;$$

$$\gamma_{ij} = \gamma_{i+1,j}; \quad i = 1,3,\ldots, \quad j = 1,2,3,\ldots \tag{5:114}$$

γ_{ij} is the interaction effect due to age interval i and for instance job j.

The job nomenclature used permits a classification both by job level and job family. In some models the interaction between job and other factors, for instance age, is reduced to an interaction between job level and age and an interaction between job family and age.

$$\gamma_{i,kj} = \theta_{ik} + \lambda_{ij}; \tag{5:115}$$

θ_{ik} is the interaction effect due to the i-th age interval and the k-th job level and λ_{ij} is the effect due to the i-th age interval and the j-th job family.

All three methods have been used in the empirical analysis. The first method is applied when data are divided into subsets of job families and each subset is analysed separately. As the estimates of the effects common to all job families, i.e. age, education, cost of living area and industry effects, turn out to be approximately equal, the results may be interpreted as if they were obtained from a simultaneous study of all job families. The second method of omitting one or more factors is in some analyses applied to the factors cost of living area and industry, and the third method is used to reduce the number of interaction parameters as described above. In all these cases one has of course to accept approximations. There is no general rule as to how large approximations are acceptable. This has to be decided from case to case. The results in section 5.2 and in appendix A show what approximations are accepted in this study.

The large-scale computations required for this kind of study sometimes give results of poor *numerical accuracy* which may become a real problem. All computations for the analysis in chapter 5 were made in double precision. To check the accuracy of the inverse matrix $(Z'Z + H'H)^{-1}$ the norm of the matrix

$$E = (Z'Z + H'H)^{-1}(Z'Z + H'H) - I; \tag{5:116}$$

i.e.

$$\|E\| = \sum_i \sum_j E_{ij}^2 ; \tag{5:117}$$

was calculated as a measure of reliability. The result obtained was usually less than 10^{-20}. (For more details about the program used see Cedheim & Klevmarken [1969].)

5.2 MODELS AND RESULTS

The results obtained when the models proposed in the previous section are applied to individual salary statistics from the Swedish Employers' Confederation (SAF) are described in this section. Because of limited computer capacity it was not possible to apply models with all possible factors and interactions on large data sets. Instead, different variants of the general linear model were applied to different subsets of data. The multiplicity of models and data sets is also explained by the application of more than one set of constraints. as well as by an attempt to use small models and data sets, when this was possible, to economize on computing expenses. In this section mostly summary measures are presented, while details of models and results are reproduced in appendix A for the most important models. The reader is referred to Klevmarken [1968a] for a complete report on the results for 1964 and previous years.

Cross section data have been used from the years 1957, 1960, 1961, 1963, 1964 and 1968 but there is no pooling of data between years. When changes in the salary structure were investigated the same model was estimated once for each year.

To investigate whether the salary structure differs between those who leave one employer and go to another and those who remain with the same employer, the data from 1960, 1961, 1963 and 1964 have been divided into the three subsets »Leavers», »Pairs» and »Beginners» which have previously been described in chapters 2 and 3.

The factors used to analyse the salary structure are age, cost of living area, education, job and industry. Others have also used for instance geographic area, plant size, race and religion. The last two factors are hardly relevant in Sweden, the first two may be relevant but they are not recorded and therefore it is not possible to investigate their effects. The analysis is restricted to male employees. The females are rather few, in particular when only employees with academic or professional high school training are considered. A more detailed survey of variables and definitions has already been given in chapter 2.

The exact variable setting of the most important models as well as the estimates are apparent from appendix A and from Klevmarken [1968a, part II]. The numbering system used in Klevmarken [1968a] is retained here. Each model is given a Roman number with possible variants named a, b, c and so forth. The data set used in a particular application of a model is indicated by B, P, L or H★ for »Beginners». »Pairs», »Leavers» and the Whole set respectively and the last two digits of the year. If necessary the constraints used are denoted

★ In Klevmarken [1968a] the following notation is used instead: B, M, S, H for Beginners (börjat), Pairs (makar), Leavers (slutat) and the Whole set (hela).

by the corresponding number. For instance, the identification II:P64, 4 stands for model II applied to data set »Pairs» 1964 with constraint No. 4. Only a brief survey of the models used is given below.

Model I includes all five factors mentioned above. Three educational qualifications are represented, namely degree in engineering, certificate in engineering I and certificate in engineering II. The most numerous technical jobs families are included. All industries are included. The model is additive. It is applied to Leavers 1963, Pairs 1963 and 1964 and to Beginners 1964 in order to evaluate differences between Leavers, Pairs and Beginners.

Models IIa–IIb are also additive and IIa only differs from model I by including employees with a degree in science and employees with »other» university training. This model is a main instrument for the analysis of the salary structure in the technical job families. It is applied to data from 1957, 1960, 1964 and 1968. For the last year some modifications of the variable set up were necessary because of a change in job nomenclature (see appendix A). IIb only differs from IIa because the job factor is omitted.

The models IIIa – IIIc are all modifications of model II. IIIa is the same as II except for cost of living area and industry which are omitted. By comparing II and IIIa it is possible to analyse the effect of this truncation of model II. To be able to introduce interactions, given a limited computer space, the technical job families are analysed in two halves. Model IIIb includes employees engaged directly in production while model IIIc includes jobs in research and development. Both models are built up of the three factors age, education and job and involve all two-factor interactions. IIIb only includes the three educational qualifications in engineering, while IIIc also includes degree in science and other university degrees. All three models are applied to Pairs 1964.

Model IV is the base for the models Va – Ve, VIa – VIj and VII which are all used in a preliminary evaluation of main effects, interactions and the consequences of omitting factors. Model IV includes all five factors but the job factor is limited to a few commercial jobs and only six industries are included. There are five educational groups, the three groups of engineers and the two economic ones. The model is additive and is applied to Pairs 1964.

Models Va – Ve are the same as IV except for one omitted factor. In Va age is omitted, in Vb cost of living area, and so forth.

The models VIa – VIj possess the same main effects as model IV but one two-factor interaction has been added to each model. Model VIa thus includes the interaction age x cost of living area, model VIb age x education and so forth.

Model VII also possesses the same main effects as IV but it also incorporates at the same time all two-factor interactions except those with industry.

The models VIIIa – VIIIg are the main instruments used for the analysis of the salary structure in the commercial, economic and administrative job families. VIIIa is additive and includes all five factors. The educational factor includes the three groups of engineers and the two groups of economists. VIIIb does not include cost of living area and industry, otherwise it does not deviate from model VIIIa. VIIIg is the same as VIIIa, except for the job factor which is omitted. Following the procedure used for the models II and III, the commercial, economic and administrative job families are also analysed in two halves. IIIc covers the commercial jobs while IIId covers the economic and administrative jobs. Both models are additive, but in VIIIe and VIIIf all two-way interactions are added to the same main effects as in VIIIc and VIIId, respectively. VIIIa is applied to Pairs 1960 and 1964 and to all 1968, while VIIIb – VIIIf are applied only to Pairs 1964.

Model IX includes all employees with any of the recorded educational qualifications. All jobs are included, but job family and job level are treated as two separate factors. Besides job and education the model also includes age, cost of living area and industry. This model gives a survey of the whole salary structure. It is estimated for 1968.

The estimation has been done with a least squares regression program described in Cedheim & Klevmarken [1969].

The analysis of the salary structure which follows starts with a few comments on the possibility of measuring salary differences in Sw.kr. rather than in relative terms and with some remarks on the stochastic properties of the models. In the next section there is an empirical evaluation of the effects of using different sets of constraints and then follows an analysis of the main effects. Possible interactions are also investigated. Main effects and interactions are ranked as to differentiating power. Some less important factors and interactions are omitted in the more detailed analysis of the salary structure in technically oriented jobs and in jobs oriented towards administration, finance and marketing. In the two final sections the differences between Pairs, Leavers and Beginners are first analysed and then changes in the salary structure from 1957 to 1968.

5.2.1 *Nominal or relative salary differences. Stochastic properties*

In chapter 4 the individual salary was made up of an initial salary to which *percentage* increases were added. This was carried over to this chapter and resulted in models where the logarithmic salary is explained by a number of dummy variables. Salary differences are then measured in relative terms. It is, however, not selfevident that a *relative* comparison is the proper one. Alternatively salaries could be compared nominally in crowns. Technically this would imply that the nominal salary, not its logarithm, is the dependent variable in our models. This alternative has been tried for instance by Hill [1959] and Holm [1970], and compared to a multiplicative model. Both Hill and Holm find that the two models give about the same fit, the multiplicative model possibly adjusts a little closer to the data, but the difference is not large enough to discriminate between the two alternatives. The same result is obtained in the present study (see Klevmarken [1968a]). Hill chooses the multiplicative model after a study of the residuals and the income distribution generated by the model. He shows that the multiplicative model generates an income distribution which resembles the actual distribution better than the nominal model. Both models generate distributions which display leptokurtosis in comparison with a corresponding normal distribution, but the multiplicative one less so. Hill shows that the leptokurtosis can be explained by the heteroscedasticity of the residuals. The residual variance of both models increases by income level, but this increase is very much stronger in the nominal model than in the multiplicative one.

A priori arguments in favour of the multiplicative model can be obtained from our knowledge about salary setting. Negotiations usually result in *percentage* increases. Demands for compensation are usually phrased in relative rather than nominal terms. In particular among salaried employees relative comparisons seem to be the habit. The same tendency is not so pronounced among workers. They very often negotiate, for instance, about additional numbers of 'öre' per hour. As a result we would expect the nominal model to be more successful when applied to wages than to salaries. In the sequel our intention to use multiplicative models will be followed.

It has been shown in chapter 4 that the individual variability increases by age. As income also increases by age we can make the same observation as Hill, that the variability increases by income. We have previously offered an explanation for this heteroscedasticity, namely that everybody does not obtain promomotion, at least not at the same rate. If now the job factor is brought into the analysis to explain salary differences, more homogeneous groups are obtained and the heteroscedasticity is reduced. Besides, the heteroscedasticity should be less of a problem in this study than in Hill's because he includes both wage earners and salary earners, while this study only embraces salary earners. There are also better possibilities of dividing the data into more narrowly defined groups than in Hill's study.

As mentioned, Hill's distribution of residual incomes (Hill [1959], table VI) is not normal, but he shows that the leptokurtic deviation from normality does not reject the possibility that the residual is normally distributed *in each factor combination* but with a varying variance, i.e. the heteroscedasticity produces the leptokurtosis in the combined distribution of residuals. Hill [1959, p.378] finds »that the separate distribution of logarithmic residuals for salary earners is closer to normality than is the overall distribution, but that for wage earners is not noticeably more normal than the joint distribution.» He also states: »It is probable that a more detailed occupational breakdown would produce distributions more closely normally distributed since there is evidently some further heteroscedasticity within the ranks of both salaried workers and wage earners.»

5.2.2 *The choice of constraints. An empirical investigation*

In section 5.1 it was shown that comparisons of effects were not in general independent of the constraints. One exception was pointed out. When the model is additive and for any of the sets of constraints 1, 2 and 4, the difference between two effects of a factor is the same for each set of constraints. It then follows that the difference between one effect defined by one set of constraints and the same effect defined by another set of constraints is the same for all effects of the factor. Three different models, all additive, have been estimated under the constraints 2 and 4. The models are named II, IV and VIIIa. The factors used are shown in table 5:4. Model II covers employees occupied in technically oriented jobs while models IV and VIIIa cover employees in commercial jobs. The models II and VIIIa are estimated on data from 1960 and 1964, and model IV only from 1964. All data used belong to the category »Pairs». Table 5:4 gives the differences in the estimates due to different constraints. For all models, in comparison with system 2, weight system 4 allocates more to the intercept and to the age factor than to education and job. The estimates of effects due to cost of living area and industry remain relatively unchanged. The estimated standard errors are less sensitive to a change in constraints than the parameter estimates. This indicates the possible danger in singling out one estimate and following the too common procedure of testing whether it deviates significantly from zero. For instance, with constraints No.2, the effect of age interval 30–34 years is estimated in model VIIIa at –0.0431 with the standard error 0.0048, but with constraints No.4 the estimate is –0.0066 and the standard error 0.0055. With the common relaxed application of the t-test, the first estimate gives the conclusion that the effect of the age interval is significantly different from zero, while the second does not give significance.

To pursue the comparison of the constraints, a model with interactions has also been estimated. Model VIh only differs from model IV by the interaction between education and job. The model is estimated on Pairs from 1964 under the constraints 2, 3 and 4. The three sets of constraints give the same result for those factors which do not interact. The differences between the three estimates of the same age effect vary from age class to age class and similarly for the job effects. However, it is perhaps more interesting to compare the three differences between two educational qualifications, two jobs and two interactions. In table 5:5 the difference between the largest and smallest effect per factor is calculated

Table 5:4. *Differences between effects estimated under the constraints 2 and 4*

Factor	Model II:P60	II:P64	IV:P64	VIIIa:P60	VIIIa:P64
Intercept	–0.009	–0.069	–0.077	–0.054	–0.028
Age	–0.046	–0.036	–0.069	–0.038	–0.052
Cost of living area	0.011	0.009	0.000	0.007	0.005
Education	0.033	0.040	0.015	0.027	0.028
Job	0.015	0.060	0.109	0.048	0.042
Industry	–0.004	–0.004	0.022	0.010	0.005
Totals	0.000	0.000	0.000	0.000	0.000

Table 5:5. *Factor ranges when model VIh:P64 is estimated under different constraints*

Factor	Constraints 2	3	4
Age	0.4293	0.4293	0.4292
Cost of living area	0.0360	0.0360	0.0360
Education	0.1530	0.1737	0.1770
Job	0.6939	0.6849	0.6801
Industry	0.1107	0.1107	0.1106
Education x job	0.1534	0.1472	0.1388

(factor range). As age, cost of living area and industry do not interact the ranges are the same for all constraints. For education, job and the interaction education x job they differ, but not by very much. At least for this model comparisons between effects do not depend very much on the constraints.

5.2.3 *Investigation of main effects*

The models used in this study involve a maximum of five factors. The question examined in this section is: Which factors are most important in the differentiation of salaries?

The results in table 5:6 are interpreted as an introduction. This table gives estimates of model IX applied to all men in 1968 with at least a (vocational) high school education. Altogether they amount to 50 120 employees. The model is additive and the factors used are displayed in the table. The constraints imposed belong to system No.2. The estimates of this model give an overall view of the salary structure in the SAF sector. Later different parts of this structure will be analysed in more detail, in particular parts defined by different sets of job families.

The estimated salary of employees in a certain age interval, working in a particular cost of living area, with a particular education and so forth is obtained by adding the age effect, the effect due to cost of living area, the educational effect and so forth to the general mean and then taking the antilogarithm of the result. For instance, the estimated salary of a 28–29 years old university graduate in engineering working on production control at job level 4 in an iron and steel plant situated in cost of living area 4 is in logarithmic form

$$\ell = 8.1362 - 0.1081 + 0.0033 + 0.1497 + 0.0338 + 0.1064 + 0.0116 = 8.3329$$

and the antilogarithm of this number is 4159, i.e. the estimated monthly salary is Sw.kr. 4159. This is not in general an unbiased estimate of the expected salary for employees with the above mentioned characteristics. Following the same arguments as in section 4.4 (p. 125), if the theoretical residual follows a normal distribution, 4159 is an estimate of the median salary of the employees belonging to the interval 28–29 years, the cost of living area 4, and so forth.

Suppose now we want to compare this salary with that obtained by those employees who have the same characteristics except for education, namely high school certificate in engineering I. Their estimated logarithmic salary is the sum of the same effects as the salary of those who graduated in engineering except for the effect due to education. From table 5:6 the logarithmic difference is then

Table 5:6. *Estimates of model IX:H68 (all men with at least high school education in 1968)*

General mean (intercept) 8.1362 (0.0006) R = 0.9229

Number of observations 50 120

Age effects

–19	–0.5184	(0.0257)
20–21	–0.3929	(0.0089)
22–23	–0.3030	(0.0038)
24–25	–0.2300	(0.0025)
26–27	–0.1669	(0.0022)
28–29	–0.1081	(0.0022)
30–31	–0.0597	(0.0021)
32–34	–0.0117	(0.0019)
35–39	0.0441	(0.0015)
40–44	0.0910	(0.0015)
45–49	0.1134	(0.0017)
50–59	0.1252	(0.0016)
60–	0.0847	(0.0036)

Cost of living area

3	–0.0282	(0.0005)
4	0.0033	(0.0014)
5	0.0648	(0.0012)

Education

University degree in		
engineering	0.1497	(0.0019)
business & economics	0.0488	(0.0046)
science	0.0710	(0.0067)
law or social science	0.0160	(0.0088)
social work	–0.1151	(0.0175)
other sciences	0.0835	(0.0061)
High school certificate in		
engineering I	–0.0022	(0.0009)
engineering II	–0.0335	(0.0008)
commerce	–0.0508	(0.0028)

Job family

0. Administration	0.0651	(0.0044)
1. Production control & supervision	0.0338	(0.0014)
2. Research & development	–0.0179	(0.0016)
3. Construction & design	–0.0290	(0.0011)
4. Other technical	–0.0327	(0.0015)
5. Journalism, library work	–0.0743	(0.0104)
6. Education	–0.0489	(0.0099)
7. General service & health	0.1613	(0.0138)
8. Commerce	0.0413	(0.0014)
9. Finance & accounting	–0.0081	(0.0027)

Job level

2	0.5345	(0.0035)
3	0.3249	(0.0018)
4	0.1064	(0.0011)
5	–0.0859	(0.0008)
6	–0.2471	(0.0016)
7	–0.3561	(0.0037)
8	–0.4015	(0.0181)

Industry

Mining	0.0090	(0.0052)
Metal and engineering industry		
Iron and steel works, metal plants	0.0116	(0.0021)
Manufacture of hardware	0.0140	(0.0034)
Engineering works	–0.0130	(0.0010)
Repair works	0.0228	(0.0070)
Shipyards	–0.0126	(0.0034)
Manufacture of electrical equipment	–0.0082	(0.0014)
Other metal industry	–0.0005	(0.0055)
Quarrying: stone, clay and glass products	0.0252	(0.0034)
Wood industry	0.0128	(0.0051)
Manufacture of pulp, paper and paper products	0.0184	(0.0030)
Printing and allied industries	0.0280	(0.0086)
Food manufacturing industries	–0.0117	(0.0045)
Beverage and tobacco industries	0.0355	(0.0130)
Textile industry	0.0046	(0.0051)
Leather, furs and rubber industries	–0.0095	(0.0005)
Chemical industry	0.0127	(0.0024)
Building and construction	0.0117	(0.0017)

$\ell_{\text{graduates}} - \ell_{\text{high school}} = 0.1497 + 0.0022 = 0.1519;$

i.e. a percentage difference between the two educational qualifications of 16.4 %.

There are several possibilities of measuring the impact of the different factors. One way is to look at the size of the estimated coefficients, but as has been pointed out already the magnitude of a particular coefficient is of relatively small interest while the salary variation due to one factor as compared with another is more interesting. One easy and natural measure is the range of the effects. Ranges have been calculated factor by factor for a number of models and tabulated in table 5:7. It is perfectly legitimate to compare the ranges factor by factor for each model, but comparisons between models should be done with great caution as the number of age intervals, jobs and educational qualifications differ between the models. For instance, the relatively small job range of model IV is explained by the lack of the low-paid job levels 6, 7 and 8 in this model.

The results in the table clearly indicate that the factor producing the greatest differentiation is the job factor. The factor ranges in the table can be transformed into percentages. For instance, the average salary in the best paid job (job level 2) is approximately 200 % higher than the average salary in the least paid job (job level 7) after differences due to other factors have been accounted for. In this sense we may say that the job factor allows for average salary differences of 200 %. It is important to note that we compare averages. The individual variability is of course greater.★ The estimates of model IX reveal that job level account for the greater part of the variability due to job while the differences between job families are not so important. In the same way we find

Table 5:7. *Salary ranges by factor*

Factor	Model and data					
	II:P60	II:P64	IV:P64	VIIIa:P60	VIIIa:P64	IX:H68
Age	0.3513	0.3088	0.3812	0.4039	0.3769	0.6436
Cost of living area	0.0823	0.0674	0.0360	0.0701	0.0525	0.0930
Education	0.1755	0.1750	0.1530	0.1713	0.1699	0.2648
Job	0.9977	0.9201	0.6929	1.0350	1.1022	
Job family						0.2356
Job level						0.9360
Industry	0.1016	0.0993	0.1111	0.1270	0.0946	0.0485

★ See section 2.2, the principle of individual salary setting. Individual merits and ability to fulfil the requirements of a job are not fully accounted for by any of factors used to analyse the salary structure.

Table 5:8. *F-ratios by factor*

Model	Omitted factor	Degrees of freedom		F-ratio	$F_{0.95}$ ★
		Numerator	Denominator		
Va	Age	4	3139	248.3	2.38
Vb	Cost of living area	1	3139	29.6	3.85
Vc	Education	3	3139	101.1	2.61
Vd	Job	3	3139	1349.0	2.61
Ve	Industry	4	3139	12.3	2.38

★ The theoretical F-values are given with 1000 degrees of freedom in the denominator.

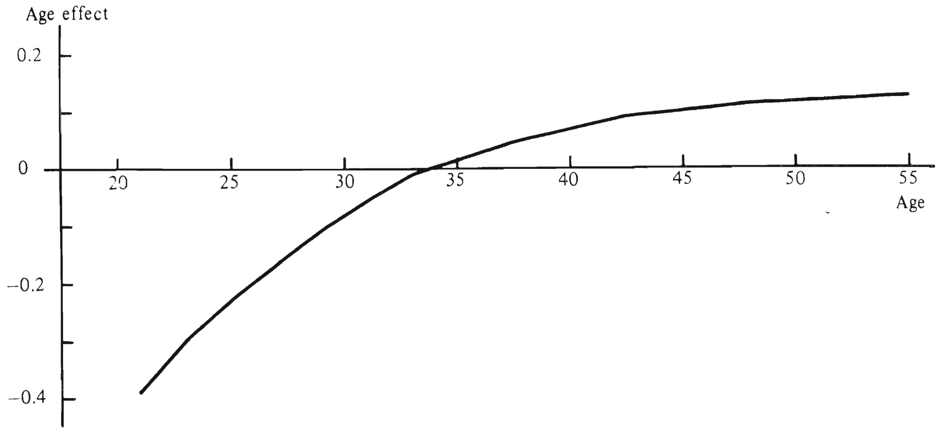
that there are average salary differences due to age of 35–50 % among the employees covered by the models II, IV and VIIIa and about 90 % when all employees are considered (model IX). The third most important factor is education which allows for differences of 15–20 % and almost 30 %, respectively. The differentiating power of the two remaining factors is considerably less, 5–10 %, and the relative importance of these two factors varies from model to model.

An alternative measure of the importance of a factor is the traditional F-ratio. As an example F-ratios have been calculated from model IV when one factor, factor by factor, is omitted. The new models are denoted Va, Vb,...,Ve. The result is given in table 5:8. Although the F-ratios are not calculated in order to test that the effects of each factor are zero, the corresponding theoretical F-values are still given in the right-hand column of the table as a base of reference. The same conclusions can be drawn from the F-ratios as from the ranges. The most important factor is definitely job. Then follows age and education and finally cost of living area and industry.

Figure 5:3 as well as table 5:6 reveal an age-salary profile of traditional form. The differences between the cost of living area, as we have seen, are relatively small. In area 5 the salaries were about 6 % higher in 1968 than in area 4 and 9 % higher than in area 3.

With the exception of graduates in social work who form a very small group in Swedish industry, academic training is paid better than high school training, but the differences are not very large. The highest average salary is obtained by graduates in engineering. They obtained in 1968 on average 7–13 % more than those who graduated in other subjects, and 15–20 % more than employees with a high school certificate. Engineers are in general better paid than economists and others. All these comparisons are standardized for differences in job attainment, and the high salaries of those with academic training who have reached this level due to quick promotion are not included in the comparisons (see below).

Figure 5:3. *Age-earnings profile of all men with at least high school education (model IX: H68)*



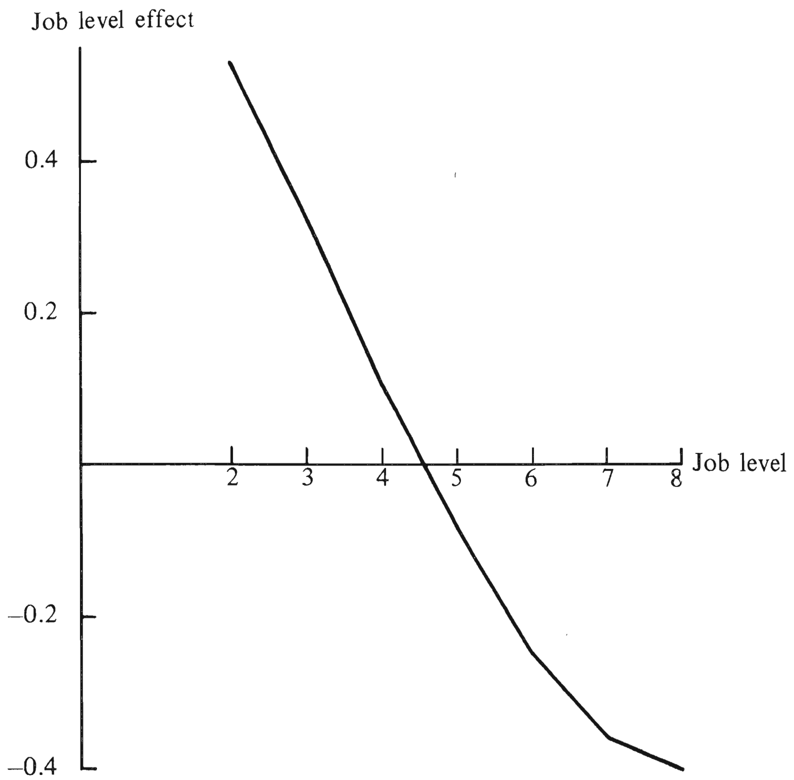
According to the estimates the best paid jobs are in general service and health. This result is due to a relatively small number of well paid physicians. Disregarding this job family, jobs in administration, production and commerce are on average better paid than jobs in research and development, construction, journalism, education and finance and accounting. The differences range from about 15 %, but between the big job families they are smaller. For instance, production work is on average paid 5 % more than research and development, and commerce also 5 % more than finance and accounting.

The job level attained by an employee who recently graduated from university is usually level 5, while those who come from a vocational high school training (in particular engineers) usually obtain level 6, and less qualified employees like typists, office clerks and card punchers obtain 7 and 8. The average salary at the top level 2 (the level below the management level) is 85 % higher than at level 5, the difference between levels 5 and 6 is 17 % and that between levels 6 and 8, 16 %. It will also be seen from figure 5:4 that the differences are smaller at the beginning of the career than in the middle and at the end.

The estimated differences between industries are small, only a few percentage units, but the standard errors are high enough to make all comparisons very uncertain.

Although the F-ratios in table 5:8 reveal a marked difference between cost of living area and industry on the one hand and age, education and job on the other, they are all significant in the sense that the calculated F-ratios are higher than the theoretical ones. We have also found that, for instance, the effects due to cost of living area are small but they are all significant, i.e. the estimates deviates from zero by more than twice their standard errors. However, this is partly

Figure 5:4. *Job level-earnings profile of all men with at least high school education*
(model IX:H68)



due to a large number of observations in each area. If they were all broken down into a number of smaller areas, some or all effects would probably become insignificant. Analogously, if insignificant effects are obtained and the sample size is then increased if possible, all effects will sooner or later become significant. The significance of the effects (factors) is thus of lesser importance then the very estimates of effect differences. This is of course not to say that the standard errors of the estimates are unimportant, on the contrary they give highly relevant information about the uncertainty of the estimates, but it is suggested that a *test* of no effect is not very meaningful. More interesting is it to *estimate* the magnitude of the effect.

In order to simplify the models to those factors most relevant to an explanation of salary differences and to avoid exceeding the memory capacity of the computer when interactions are added to the models, it is desirable to omit one or two factors. Cost of living area and industry have been found to have very much less differentiating power than the three other factors. It is therefore natural to omit these two variables, but before doing so an investigation should be made as to the effect this truncation of the models will have on the estimates of the age, education and job effects. If the regressors belonging to cost of living area and industry are orthogonal to all other regressors, then the truncation does not change the estimates (see section 5.1). Although the age distribution, the distribution of educational qualifications and jobs do not differ very much from one cost of living area to another and from one industry to another, there are certainly differences (see chapter 3). Instead of investigating the intercorrelations between regressors it is empirically more convenient to measure the result of omitting one factor directly in the estimates of the remaining effects. This will be done below, first with model IV and then with II and VIIIa. To begin with only one factor is omitted. Table 5:9 shows the differences between the estimates of model IV:P64 and the estimates of the models Va – Ve:P64. As before model Va is the same as model IV except that the age factor is omitted. In model Vb cost of living area is omitted, and so forth. Omitting cost of living area or industry hardly changes the estimates of the age, education and job effects at all. No change exceeds one per cent. The change obtained when both cost of living area and industry are omitted from model IV is not investigated but will be when they are omitted from models II and VIIIa. The truncated models are called IIIa and VIIIb respectively. Table 5:10 shows the differences between the estimates of II and IIIa. Dropping both cost of living area and industry does not produce changes in the estimates for age and education worthy of consideration, but some of the job estimates change more than should be acceptable. The table reveals that there are some relatively high negative differences for production and supervision jobs (110, 120, 130). A possible explanation is that industries paying relatively high salaries also have a large proportion employed in these jobs. The industry effects are then picked up by the job effects. The industry estimates of model II:P64 show that employers, particularly in Building and construction, pay more than other employers but also Iron and steel works; Metal plants; Quarrying: stone, clay and glass industries; Repair works; Pulp and paper industry and Mining pay salaries a little above average. As has been pointed out already the differences are small and the standard errors relatively high. From chapter 3 we know that there are many engineers in Building and construction employed in jobs belonging to the families 120 and 130.

The comparison between models VIIIa and VIIIb reveals the same insensitivity to the omitted factors as before (see Klevmarken [1968a] p.43) and although this result formally only applies to the particular models investigated, it can probably be extended to the whole SAF sector.

Omitting other factors than cost of living area and industry may teach us more about the interrelations between the factors. In particular it may be interesting to omit the job factor because education more or less predestinates to certain jobs, not only to job families but also to job levels. In chapter 3 it was demonstrated that employees with academic training were in general employed at higher job levels than employees with less training and engineers are generally found at higher levels than employees with training in business and commerce. In chapter 4 age-earnings profiles were studied. The age dependent salary increments estimated to generate these profiles can at least partly be explained by promotion from one job level to another and the differences observed between the educational qualifications can possibly be explained by differences in promotion.

Table 5:9 shows that omitting any of the factors age, education or job has a considerable effect on the other two factors. The last-but-one column of the table shows that the effects of the omitted job factor are first picked up by the age factor but also by education. The age differences increase because the high salaries at the top levels, usually obtained by middle aged and old employees, are now partly measured as an effect of age and similarly for the low salaries obtained by younger employees at low job levels. The job effect also spills over to the educational effects. The effects due to university training in engineering and in business & economics increase while the effects due to high school training in engineering decrease. The explanation is of course that graduates fill most of the top level jobs while non-graduates are more frequent at the low levels. For the same reason the omitted educational effects in the third column show up as an increased differentiation due to job. As shown in chapter 3 the relative frequency of graduates is relatively high in the age interval 23–35 years while in particular non-graduate engineers II are frequently middle-aged. This explains the slightly higher age effects in the intervals 26–29 and 30–34 years when education is omitted, and the higher educational effect due to engineering II and the lower effects due to university training in engineering or business & economics.

In figure 5:5 age-earnings profiles have been plotted based on estimates of model II for 1968. The solid curve is the profile obtained when the model includes the job factor, while the broken curve is the profile obtained without it. The steeper broken profile shows how much promotion means as regards salary differences between young and old employees. The difference between the curves can be taken as a measure of the effect of promotion. The solid curve would then indicate salary differences due to age regardless of promotion. Figure 5:6 shows the same two profiles for the commercially oriented jobs of model VIIIa for 1968. Although the two models cover entirely different job families and partly different educational qualifications the similarity between the profiles in figures 5:5 and 5:6 is remarkable.

Table 5:9. *Differences between the estimates of model IV and the models Va – Ve; Pairs in 1964* ★

Factor-effect	Models/Factors omitted				
	IV–Va Age	IV–Vb Cost of living area	IV–Vc Education	IV–Vd Job	IV–Ve Industry
Intercept	–0.0001	–0.0001	–0.0001	–0.0001	–0.0001
Age					
–25		–0.0002	0.0001	0.1801	–0.0035
26–29		–0.0001	–0.0118	0.1584	0.0000
30–34		0.0016	–0.0107	0.0987	–0.0009
35–44		–0.0004	0.0034	–0.0180	–0.0001
45–59		–0.0002	0.0080	–0.0972	0.0009
60–		–0.0001	–0.0037	–0.1104	0.0011
Cost of living area					
3	–0.0015		–0.0022	0.0051	–0.0022
4	0.0065		0.0059	–0.0105	0.0008
5	–0.0010		0.0003	–0.0029	0.0032
Education					
Graduate in engineering	0.0147	–0.0003		–0.1312	0.0021
Certificate in engineering I	0.0037	0.0008		0.0262	0.0024
Certificate in engineering II	–0.0208	–0.0026		0.0628	0.0005
Graduate in business & economics	0.0582	0.0020		–0.1742	–0.0060
Certificate in commerce	0.0052	0.0047		–0.0029	–0.0070
Job					
8102	–0.0774	0.0015	–0.0451		–0.0004
8103	–0.0557	–0.0018	–0.0247		–0.0024
8104	–0.0226	0.0009	–0.0061		0.0008
8105	0.0523	–0.0002	0.0219		0.0004
9454	0.0183	0.0023	–0.0053		0.0032
Industry					
Metal and engineering industry	0.0001	–0.0008	–0.0028	0.0020	
Quarrying: stone, clay and glass industries	0.0103	0.0054	0.0131	0.0021	
Manufacture of pulp, paper and paper products	–0.0110	0.0030	0.0247	–0.0060	
Textile industry	–0.0066	0.0094	–0.0079	–0.0264	
Chemical industry	0.0143	0.0028	0.0093	–0.0379	
Building and construction	–0.0147	0.0010	0.0096	0.0247	

★ Constraints No. 2 have been used.

Table 5:10. *Differences between the estimates of model II and model IIIa; Pairs in 1964**

Intercept	0.000		
Age		*Education*	
–25	0.0012	Degree in engineering	–0.0051
26–29	–0.0002	Certificate in engineering I	0.0062
30–34	–0.0016	Certificate in engineering II	–0.0028
35–44	–0.0003	Degree in science	–0.0007
45–59	0.0013	Other academic degrees	0.0035
60–	0.0006		

Job

Job family	Job level 2	3	4	5	6	7
110	–0.0074	–0.0121	–0.0055	–0.0112		
120			–0.0752	–0.0701		
130				–0.0105	–0.0733	–0.0719
200	0.0279	0.0267	0.0232	0.0095	–0.0246	
210	0.0085	0.0068	0.0061	0.0073	0.0031	–0.0109
310	0.0124	0.0097	0.0094	0.0084	0.0116	–0.0017
400	0.0110	0.0055	0.0070	0.0096	0.0147	
470		0.0053	0.0008	0.0050	0.0084	
490	0.0052	0.0053	0.0024	0.0054		

* Constraints No. 2 have been used.

Figure 5:5. *Age-salary profiles from model II: H68, with and without the job factor*

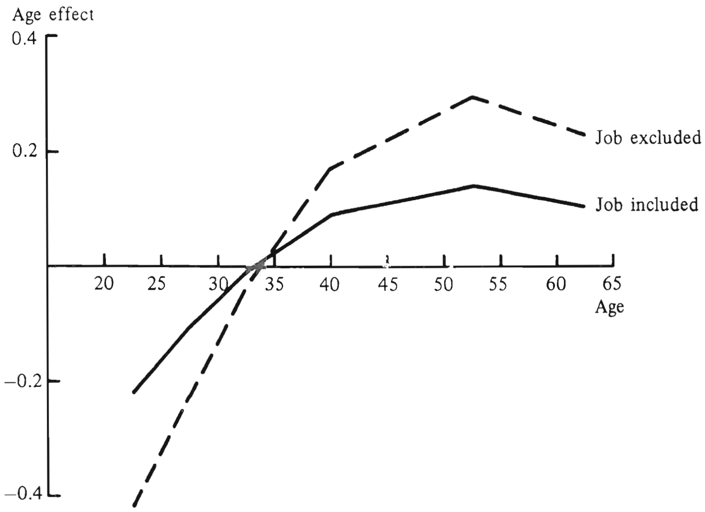
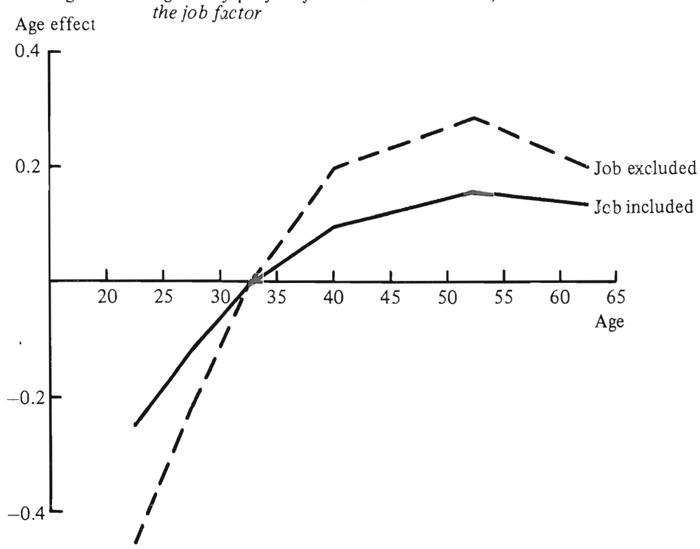


Figure 5:6. *Age-salary profiles from model VIIIa: H68, with and without the job factor*



These profiles can be used to calculate »increments per year» more or less comparable with the γ-values in chapter 4. Disregarding the fact that the profiles in the two figures are standardized for differences in cost of living area and industry and that the profiles in chapter 4 were calculated for each education, the broken profiles are comparable to the cross section profiles in chapter 4. In table 5:11 the logarithmic salary differences between successive age intervals have been recalculated to annual changes. If the average increase in initial salaries, 0.06 -- 0.07, is added to the numbers obtained when job is omitted, the result is roughly comparable to the γ-values in chapter 4. The difference between the columns »Job excluded» and »Job included» indicates how much promotion contributes to the annual salary increases, between 1.5 % and 2 % per year below (approximately) 40 years and almost nothing above. Model II and model VIIIa give about the same result. As these calculations have been made from a cross section and only from one cross section the interpretation given cannot be accepted without reservations. It has already been proposed in chapter 4 that the promotion activity is most intense during periods of excess demand for labour. It must also depend on the age distribution of the labour force. To use one cross section may then give a »non-typical» result. More serious is perhaps the fact that the profiles are assumed to be the same for graduates and non-graduates. Graduates enter the labour market later than non-graduates and they probably obtain higher increases due to promotion when they are in their mid-thirties than non-graduates. The result may very well be that we underestimate the increase due to promotion for both groups.

Table 5:11. *Age effect differences (per year) estimated from models II:H68 and VIIIa:H68, with and without the job factor*

Age interval	Logarithmic increase per year			
	Model II		Model VIIIa	
	Job excluded	Job included	Job excluded	Job included
22.5–27.5	0.038	0.023	0.047	0.026
27.5–32.5	0.039	0.019	0.042	0.023
32.5–40.0	0.028	0.013	0.027	0.013
40.0–52.5	0.010	0.004	0.007	0.005
52.5–62.5	–0.007	–0.003	–0.009	–0.002

The changes in the educational effects when the job factor is omitted from Model II and model VIIIa are demonstrated in figure 5:7. In both models the differences between graduates and non-graduates are increased. In the technical job families (model II) the effect of graduate engineering is increased by 11 percentage units and the effect of engineering II decreased by 15 percentage units. In the commercial and accounting families (model VIIIa) the effect due to graduate studies in business & economics increases by 14 percentage units which is more than the 10 percentage units for graduate engineers. In these job families promotion therefore means more to businessmen and economists than to engineers, but the latter are still receiving higher rewards. The effects of engineering II and commerce both decrease by 9–10 percentage units. These numbers then indicate how much promotion means as regards the salary differentiation between educational qualifications. For instance, a graduate in business & economics receives on average a salary which is 8 % higher than that received by a high school commercially trained employee (model VIIIa) when differences in job attainment are accounted for. When there is no standardization for job differences the graduate obtains on average 35 % more than the non-graduate. The difference, 27 %, can roughly be interpreted as the effect of the better promotion a university graduate obtains.

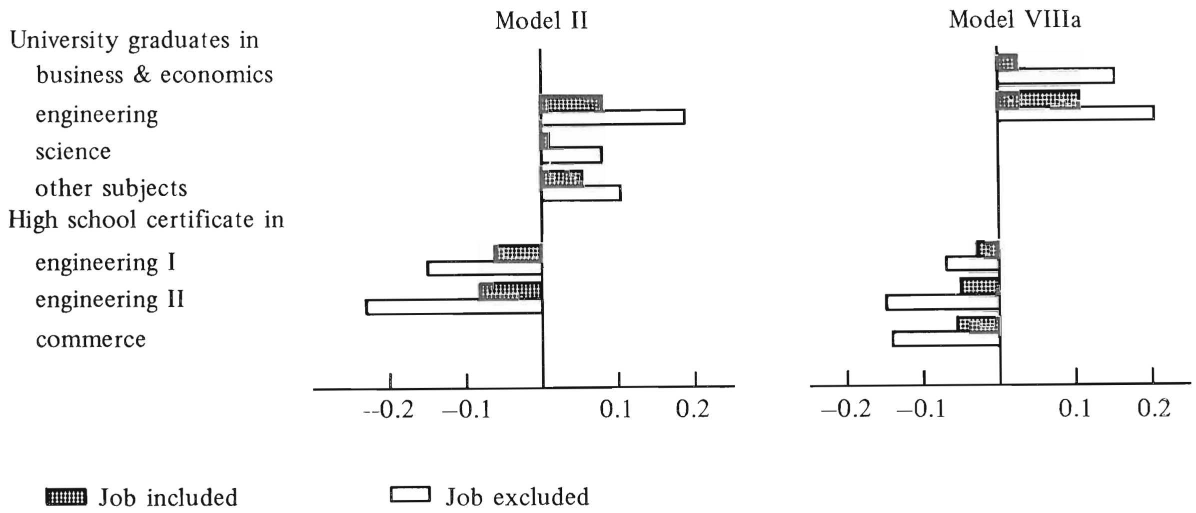
5.2.4 *Investigation of interactions*

In the previous section all models investigated were additive. Now is the time to introduce interactions and to compare the magnitude of these with the main effects. First some a priori arguments for the existence of interactions and their directions.

From other studies we know that labour with an advanced education and employed at top level jobs shows a higher geographical mobility than do others (Rundblad [1964], p.173). On a local level the supply of jobs which meet their specialization is limited. They therefore look for employment in a geographically wider labour market. This kind of labour is usually better informed about employment, salaries, etc. than less educated labour which reduces geographically salary differences between them because of market imperfections. The experience of Swedish employers shows that they usually have to pay highly qualified labour at least as much in rural areas as in the big cities. From this we may infer that salary differences between cost of living areas should be less among graduates and at high job levels than among non-graduates at low job levels.

It is also well known that young people have a higher mobility, including geographical mobility, than middle aged and old people (see section 3.4 and

Figure 5:7. *Educational effects of the models II and VIIIa, with and without the job factors in 1968*



Rundblad [1964], p.162). For the same reason as above we would then expect a tendency to smaller salary differences between cost of living areas for young than for old employees.

The estimated profiles in chapter 4 revealed an increased salary difference by age between educational qualifications in particular between graduates and non-graduates. This result thus shows an interaction between age and education, but no allowance was made for differences in job attainment. The introduction of job level should at least reduce this interaction.

Formal education is generally considered to be less important at high job levels, partly because these jobs are often of a rather general managerial nature and partly because these employees, who are relatively old, have a more or less obsolete formal education. Salary differences between educational groups may then be smaller at high job levels than at low levels. It is also obvious that formal education is more needed in some job families than in others. For instance, some expert scientific and technical knowledge and skill is usually required for research which is not always very competitive in other fields. This then suggests the presence of interaction between education and job family. However the salaries paid to these experts naturally depend on the situation in their rather limited markets which could neutralize the interaction we would normally expect to find between education and job family.

Interactions with industry are perhaps more difficult to predict. From chapter 3 we know that some educational qualifications and some jobs are more frequent in some industries than in others. This indication of differences in relative advantage may then also be an indication of possible interactions. The estimates of the main effects were, however, relatively small and uncertain and there are no reasons to expect anything else for the interactions, although exceptions may be discovered.

When analysing British income data Hill [1959] used a model with nine occupations, six age groups, four regions, three town sizes and ten industries. Some of the variables were combined into interactions, namely occupation x age, occupation x region and occupation x town size. He found that income differences between young and old were greater among those employed in managerial and professional occupations than among manual workers. The difference between the age groups 18–24 years and 55–64 years was 66 % for managerial and higher professional occupations and only 8 % for manual workers. The interaction between occupation and region was smaller. For instance, the difference between Midland & Wales and Scotland was 7 % in the managerial and professional occupations and virtually none for shopkeepers and unskilled manual labour. Although small, this interaction is opposite to what may be expected for Sweden. The possible explanation for this result is that it is standardized for differences in town size. The income differences in the United Kingdom between conurbations and rural areas are greater for shopkeepers and unskilled manual workers than for the managerial and professional occupations, but still rather small, 6 % and 2 %, respectively. Unfortunately Hill did not calculate any standard errors of the estimates which can be used to judge the uncertainty in these comparisons.

Some results from Holm's study of males and females in Sweden who were employed full time over the whole year (SOU [1970:34]) may also be of some interest. His model includes the factors age, education, branch of economy and region. The only interactions which give some contribution towards the explanation of income differences are for males age x education and education x branch of economy and for females education x region, education x branch of economy and age x branch of economy. The difference between the age groups 25–34 years and over 55 years is 7 % for males with secondary school education with or without vocational training, while for males with a university degree the difference is 45 %. To make a correct interpretation of this interaction it must be kept in mind that the model does not include any occupational factor. The difference between income earners with a low and a high education, respectively, is at least partly due to different occupational mixes. Holm's estimates of the interaction education x branch of economy seem to be a little erratic with high standard errors. For instance in agriculture, forestry, hunting and fishing males with a degree have 114 % more income than males with high school education (with or without vocational training), while there appear to be nearly no differences in public service! The different set-up of interactions for females is also a little puzzling.

In a first attempt to determine how important interactions are for the explanation of salary differences in Swedish industry, two-way interactions have been added one by one to model IV to form the new models VIa – VIj. Model VIa contains the interaction age x cost of living area, VIb age x education, and so forth. The differentiating power is measured in the same way as before by ranges and F-ratios. The ranges have to be used with some caution because cross classification sometimes reduces the number of observations in each combination to just a few. In this sense some estimates are obtained from a small number of observations and their estimated standard errors become fairly large. To prevent too much stochastic variation in the ranges, they have only been calculated on estimates from at least five observations.

Table 5:12 shows the calculated ranges and F-ratios. They do not exactly tell the same story. The ranges do not differentiate very much between the interactions. The magnitudes indicate that the interactions are of the same importance as regards the explanation of salary differences as the main effects education and industry (table 5:7). Cost of living area x industry, job x industry and age x industry are the interactions with the widest range. The F-ratios are very small compared to the F-ratios of the main effects. Two ratios, a little higher than the rest, can be distinguished, namely for cost of living area x industry and age x job. Table 5:13 shows the estimated interactions of cost of living area x industry and the corresponding number of observations. Although this is one of the two interactions which contributes most to the explanation of the salary differences, inspection of the estimates and the corresponding standard errors reveals that most estimates are very uncertain. Remembering this we may notice that the estimated salary differences between cost of living areas 5 and 3 are more than 15 % greater than average in Quarrying: stone, clay and glass industries, while in the Textile industry the difference between areas 4 and 3 is 13 % less than average. In Building and construction the salaries are 7 % lower than average in area 4 and 5 % higher than average in area 5.

The weakness of the analysis of models VIa – VIj is that only one interaction is treated in each model. The estimates of for instance age x job may have picked up effects from other interactions now left outside the model. To investigate this and in an attempt to obtain less uncertain estimates, model VII was estimated. This model is obtained from model IV by adding all two-factor interactions except those with industry. In forming interactions the age groups are reduced to three, while the six original groups are kept for the main effects. The adjusted ranges have been calculated for table 5:14. They are now reduced to a magnitude even smaller than before which is of course at least partly due to the reduced number of age groups. Two ranges are a little wider than the others, education x job and age x job. Inspection of the first group of estimates reveals that the relatively wide range entirely depends on one estimate from nine observations only. If the range is recalculated with this estimate omitted it is reduced to 0.0715. The interaction age x job is illustrated in figure 5:8★ which clearly shows how the salary differences between age groups increase at high job levels. A result in close agreement with those obtained by Hill and Holm.

★ Job 9454 is omitted from the figure.

Table 5:12. *Adjusted ranges and F-ratios for interactions in the models VIa–VIj*

Model/Interaction	Adjusted range	F-ratio	Degrees of freedom in numerator	$F_{0.05}$
VIa age x cost of living area	0.1137	2.00	10	1.84
VIb age x education	0.1261	0.69	20	1.58
VIc age x job	0.1316	4.11	15	1.68
VId age x industry	0.1687	2.04	25	1.52
VIe cost of living area x education	0.0759	2.35	8	1.95
VIf cost of living area x job	0.0876	1.24	8	1.95
VIg cost of living area x industry	0.2179	5.93	9	1.95
VIh education x job	0.1534	1.66	16	1.65
VIi education x industry	0.1467	1.76	20	1.58
VIj job x industry	0.2066	2.61	18	1.62

Note: The adjusted ranges are only calculated on estimates from at least five observations.

The F-ratios are »F to enter» interactions into model IV and the degrees of freedom in the denominator are 3139. The theoretical F-values are, however, obtained for 1000 degrees of freedom in the denominator.

Table 5:13. *Interactions cost of living area* x *industry in model Vg:P64*

Industry	Estimates of interaction effects			Number of observations		
	Cost of living area					
	3	4	5	3	4	5
Metal and engineering industry	0.002 (0.001)	0.004 (0.003)	−0.005 (0.002)	1289	406	855
Quarrying: stone, clay and glass industries	−0.042 (0.010)	0.048 (0.028)	0.119 (0.032)	78	23	18
Manufacture of pulp, paper and paper products	−0.007 (0.009)	0.036 (0.027)	−0.022 (0.041)	81	24	12
Textile industry	0.037 (0.013)	−0.099 (0.034)	–	37	14	0
Chemical industry	−0.008 (0.008)	0.056 (0.026)	−0.018 (0.024)	115	26	30
Building and construction	0.020 (0.019)	−0.067 (0.016)	0.053 (0.016)	44	55	53

Figure 5:8. *Interactions age x job in model VIII:P64*

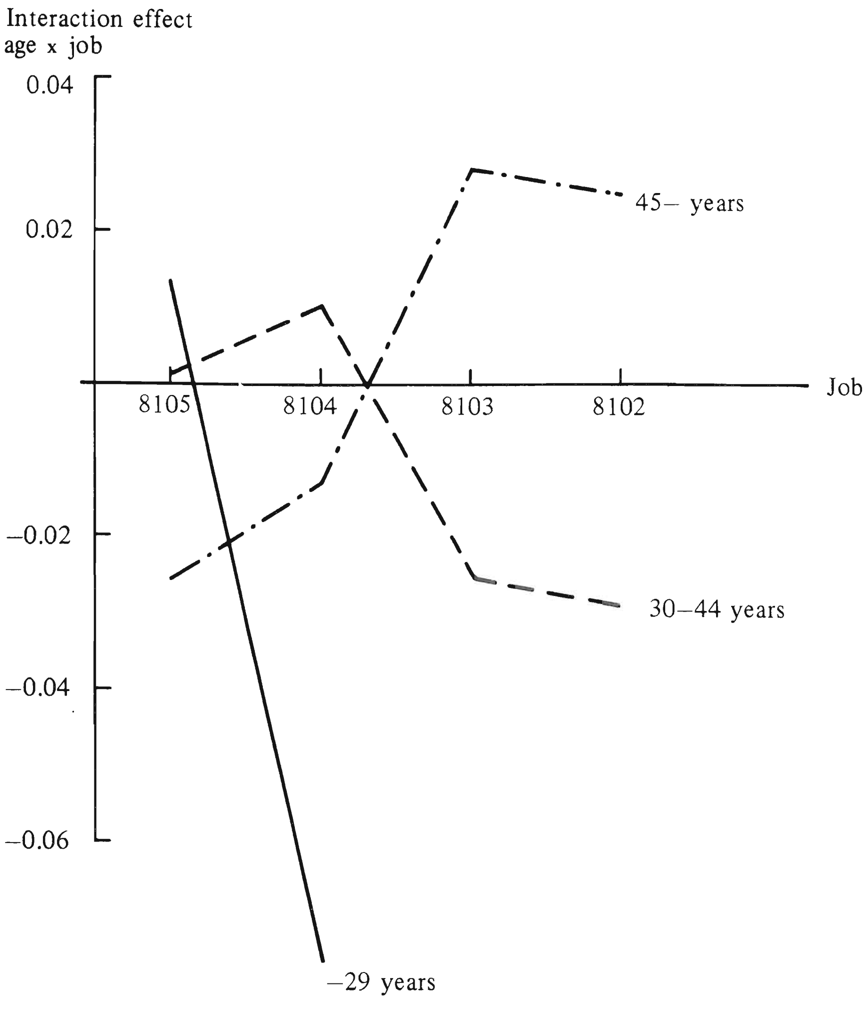


Figure 5:9 shows that the salary differences between the cost of living areas, in particular between area 5 on the one hand and areas 4 and 3 on the other, are higher for non-graduates than for graduates. In area 5 non-graduates are paid about 3 % more than in area 3, while in area 5 graduates are paid about 1 % less than in area 3. This pattern is in accordance with our previous expectations.

In figure 5:10 the cost of living areas are compared by age. The cost of living area effects and the interactions are plotted but not the age effects. There are almost no differences between the areas among employees older than 44 years but a difference of approximately 6 % between area 5 and areas 3 and 4 for employees younger than 30 years. In this case there are apparently forces stronger than mobility, as the result is opposite to the predicted result.

All these comparisons are very uncertain and they serve mainly an illustrative purpose. Because of the uncertainty it is of no great interest to pursue the comparisons. The small contribution of the interactions is also illustrated by the F-ratio in table 5:15, as compared to those of the main effects. Model VII is limited to a relatively small set of data and the results obtained may not be valid in a more general context. Interactions will therefore be studied also in the two following sections which are used for a study of the salary structure in technical jobs and in commercial, financial and accounting jobs, respectively. The two F-ratios for model VIII in table 5:15 are, however, already an indication on the relatively small importance of interactions in the job families commerce, finance and accounting.

Figure 5:9. *Comparison between cost of living areas by education in model VII: P64*

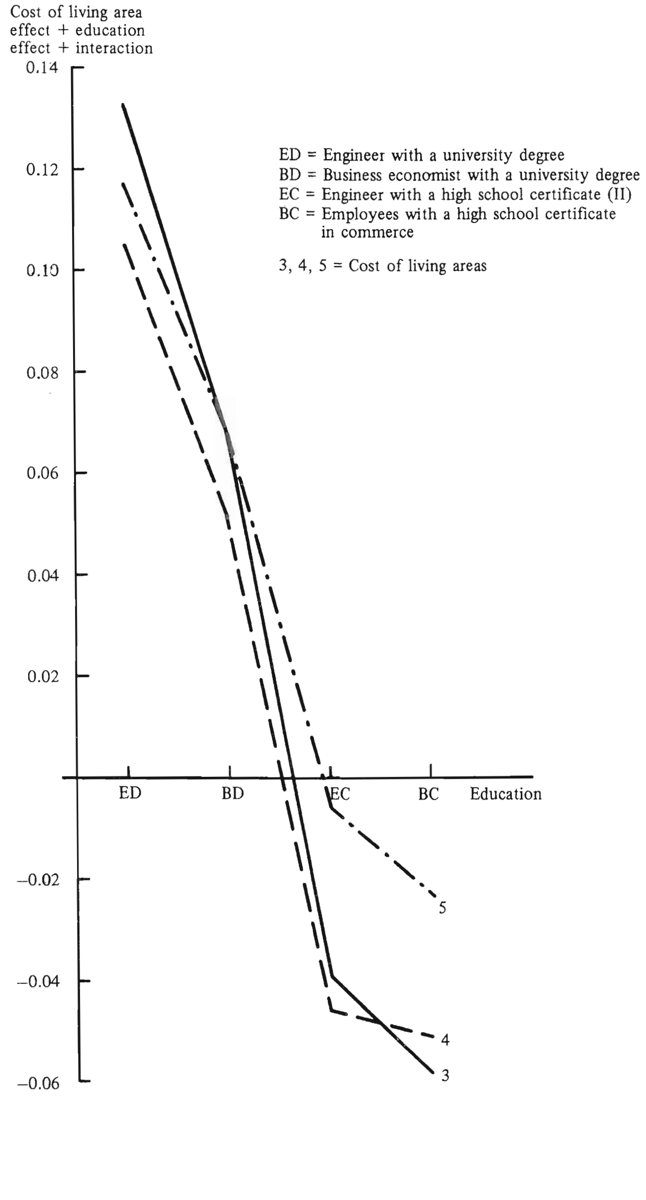
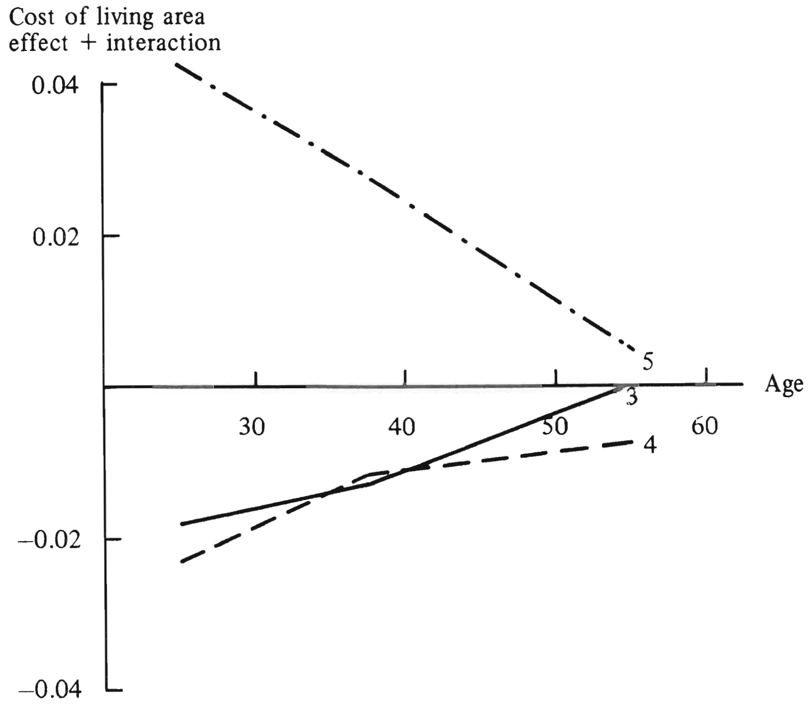


Figure 5:10. *Comparison between cost of living areas by age in model VII: P64*



3, 4, 5 = Cost of living areas

Table 5:14. *Adjusted ranges of interactions in model VII*

Interaction	Adjusted range
Age x cost of living area	0.0382
Age x education	0.0504
Age x job	0.1042
Cost of living area x education	0.0840
Cost of living area x job	0.0648
Education x job	0.1713

Table 5:15. *F-ratios for interactions*

Models with and without interactions	F-ratio	Degrees of freedom		$F_{60,\infty}(0.05)$
		Numerator	Denominator	
VII/IV	2.18	50	3139	1.32
VIIIe/VIIIc	3.15	54	5141	1.32
VIIIf/VIIId	3.60	52	2347	1.32

5.2.5 *The salary structure in some important jobs belonging to the technical, economic and administrative job families*

The reader has now been shown the general magnitudes of main effects and interactions. This section is devoted to a study in some detail of the salary differentiation in the most numerous technical, economic and administrative jobs. The technical job families are analysed separately from the economic and administrative families and the results are then compared. In order to neutralize differences in age distribution, distribution by cost of living area, education and industry, constraints No. 4 are sometimes used.

Model II is an additive model which covers the technical job families, while model VIIIa which is also additive covers the non-technical families. Both models include the five factors age, cost of living area, education, job and industry. Although the factors are called the same, their composition differs a little. In addition to the job factors which of course have no effect in common, the educational set-up also differs. Both models include the three categories of engineers but model II in addition includes employees who have graduated in science and employees with other university training, and model VIIIa employees with a degree in business & economics and with a certificate in commerce. The data set used is Pairs from 1964. The salary structure is now analysed for this year only. An investigation of the permanence of the structure is postponed to a following section.

A comparison between the two intercepts shows that the general salary level in the technical job families is very nearly the same as the level in commercial, economic and administrative work. The two age profiles are drawn in figure 5:11, and the small difference between the two curves reveals somewhat wider salary differences between younger and middle aged employees in the non-technical jobs than in the technical ones. Cost of living area differentiates a little more in model VIIIa than in II (figure 5:12). On average salaries in area 5 are 7 % and 5 %, respectively, higher than in area 3. The salary differences between the three groups of engineers are roughly the same in technical and non-technical jobs. University engineers obtain on average a considerably higher salary than other academics and non- academics. They obtain about 7 % more than business economists, 8–9 % more than scientists and other academics and 15–18 % more than non-academics.

The job estimates first of all reveal the great differences between job levels but there are also systematic differences between job families. Figures 5:14 and 5:15 illustrate the differences between a few big job families from each group. Employees engaged directly in production receive a higher reward than those who are not. Employees in research and development thus earn about

Figure 5:11. *Age-earnings profiles for the technical job families* (model II: P64,3) *and for the families commerce, finance and accounting* (model VIIIa: P64,3)

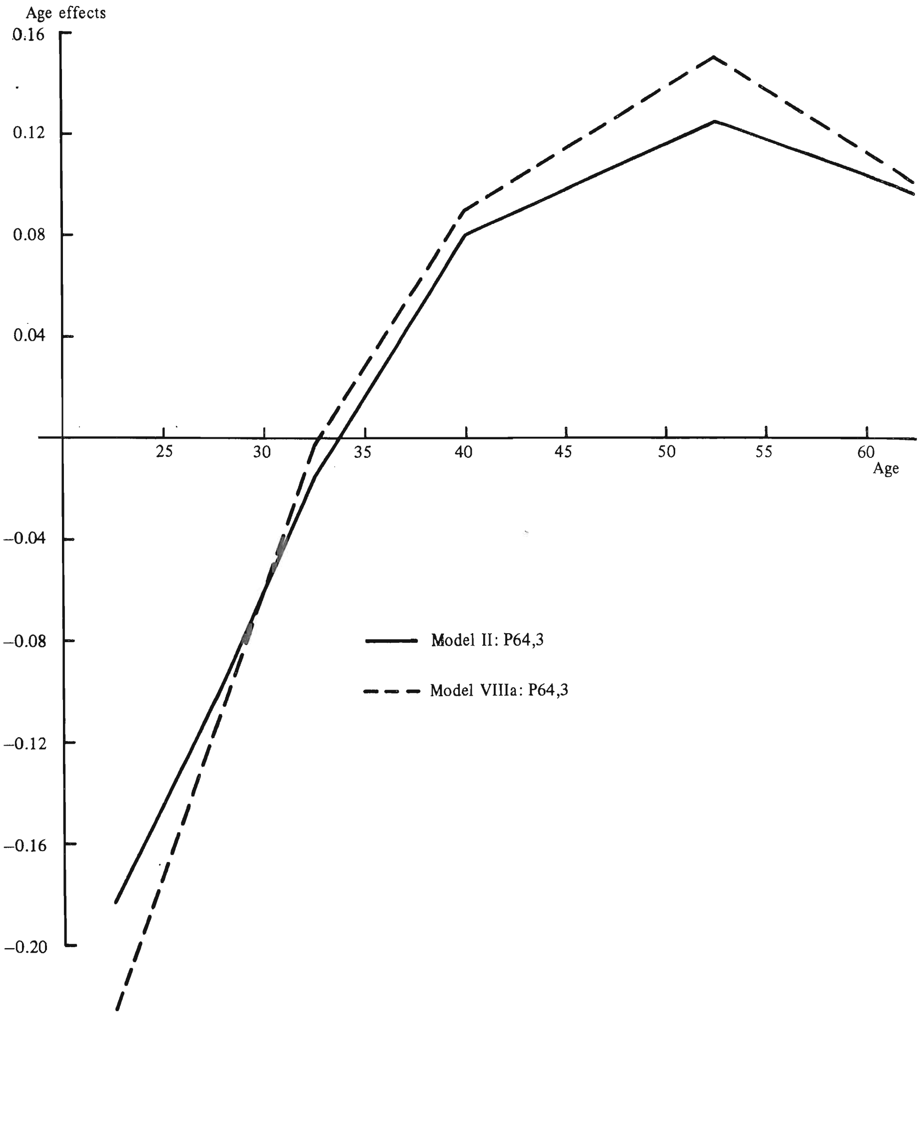


Figure 5:12. *Salary differences between cost of living areas in technical jobs* (model II: P64,3) *and in commerce, finance and accounting* (model VIII: P64,3)

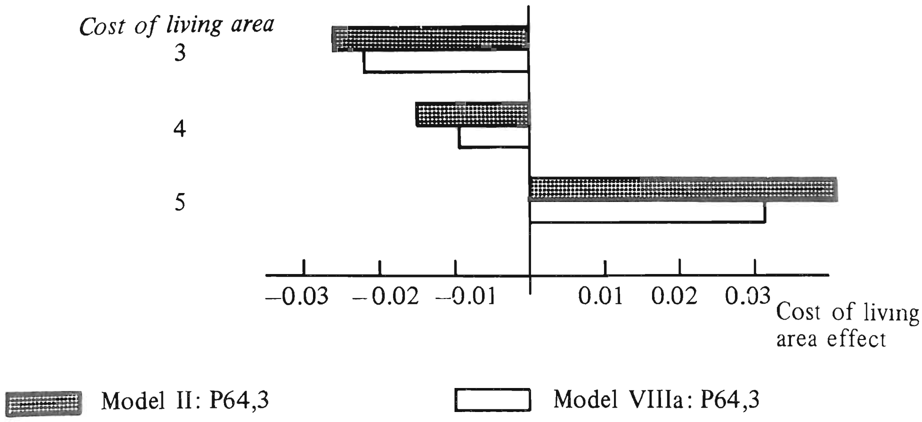


Figure 5:13. *Salary differences between educational qualifications in technical jobs* (model II:P64,3) *and in commerce, finance and accounting* (model VIIIa:P64,3)

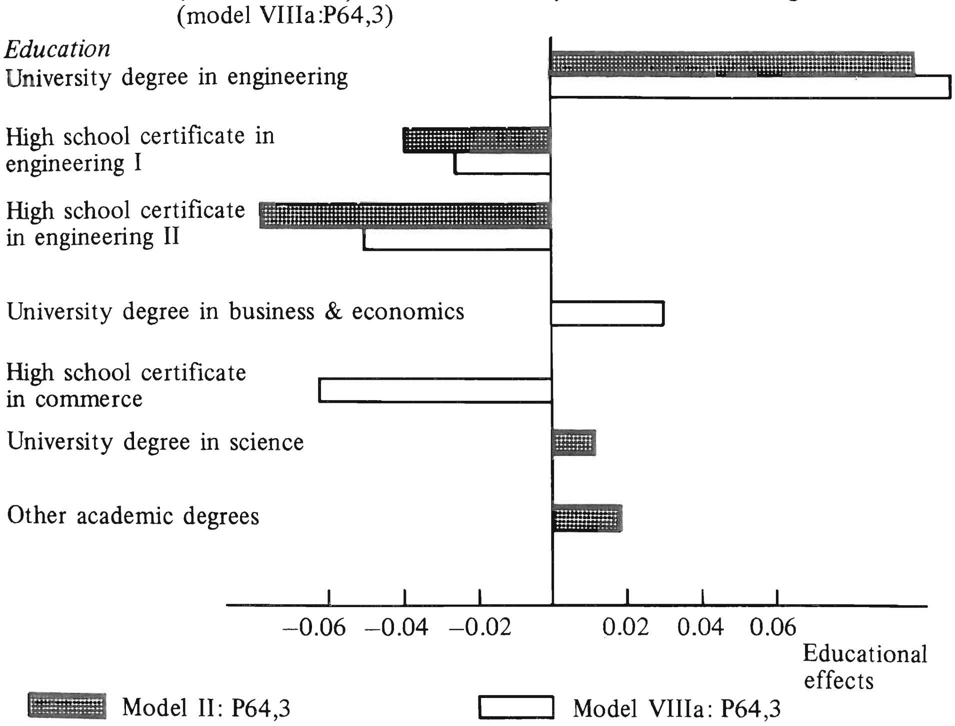


Figure 5:14. *Job effects in technical work* (model II:P64,3)

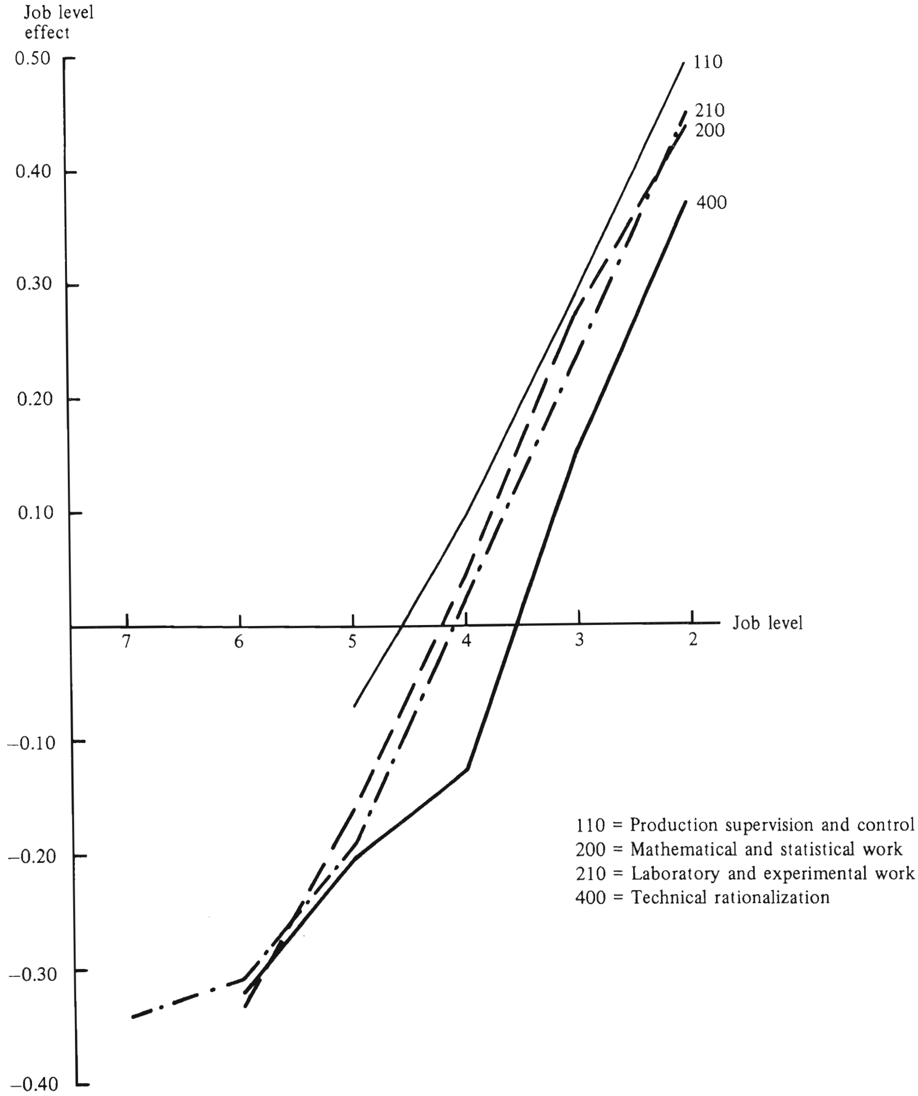
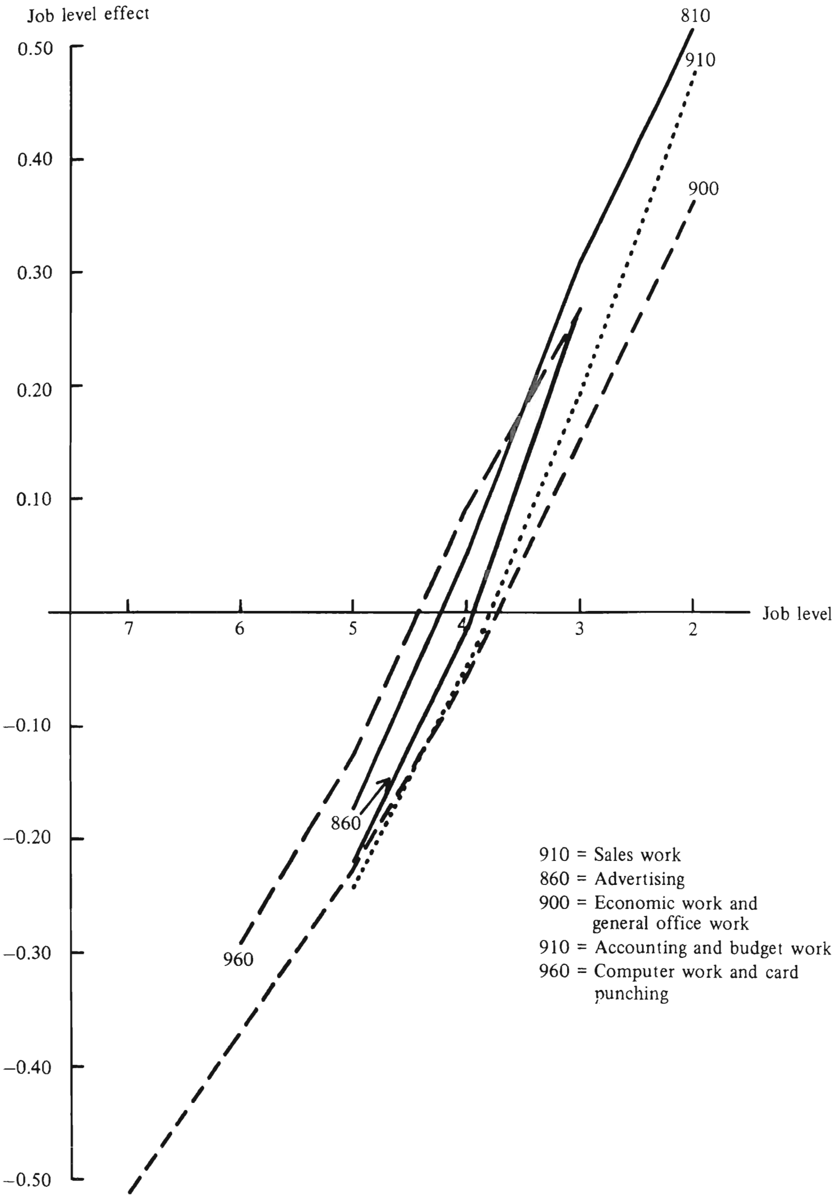


Figure 5:15. *Job effects in commercial, economic and administrative jobs* (model VIIIa: P64,3)



5 % less than those in production. The difference is even greater at job level 5. Those who work on time study and technical rationalization belong to a job family with a low salary level. On the commercial, economic and administrative side there is a similar differentiation. Sales work is paid equally well as productive work on the technical side. At the lower job levels computer work and card punching are paid even better. Accounting and budget work, other economic work and general office work are paid less. In particular the difference between a sales manager and an officer manager is more than 15 %.

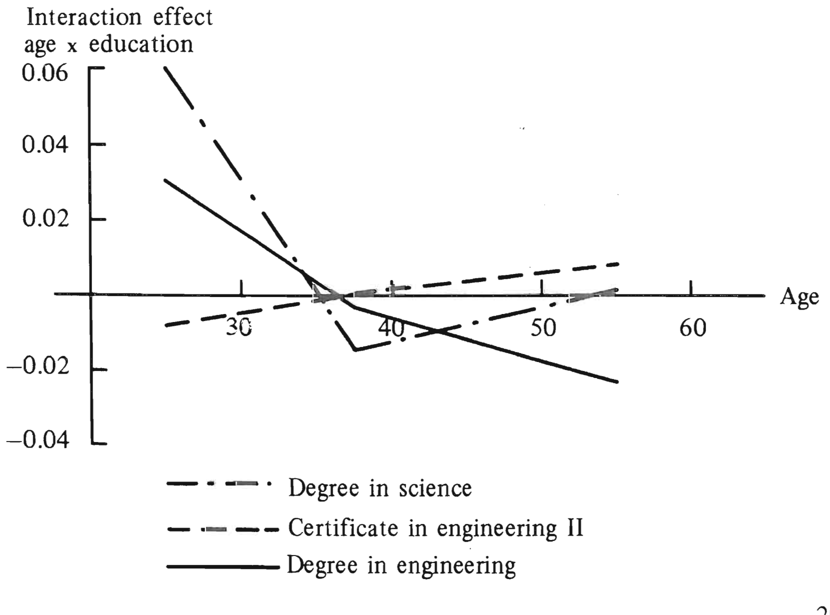
Except for Building and construction, the differences between industries are typically less than 2–3 % in technical jobs. In Building and construction salaries are 7 % higher than average. Quarrying: stone, clay and glass industries which pay nearly 2 % more than average are also influenced by the »building boom». Pulp and paper industry and Iron and metal works also belong to the industries which paid somewhat more than average in 1964. Among others Food manufacturing industries; Printing industry; Textile industry; Engineering works and Shipyards paid 1–2 % less than average. For Commercial, economic and administrative jobs also Building and construction and Quarrying: stone, clay and glass industries paid more than average, but only about 3 % more. Manufacture of electrical equipment; Other metal industry; Printing industry and Food manufacturing industries paid about 3 % less than average.

Because of limited computer capacity it was not possible to add all possible two-factor interactions to the additive models II and VIIIa. To save space cost of living area and industry were omitted from the models and only the interactions between the remaining factors age, education and job were estimated. As is already shown in section 5.2.3 the estimates of age, education and job effects hardly change at all when cost of living area and industry are left out. The interpretation of the effects of these two factors can safely be done from the additive models. In order to reduce the size of the models further they have each been divided into two submodels. Model IIIb only includes employees engaged directly in production while model IIIc covers research and development. In a similar way model VIIIe covers the commercial jobs, while VIIIf covers accounting, administration and general office work. As the jobs now included in the models IIIb and VIIIe are on average better paid than the jobs in IIIc and VIIIf, the intercept increases and all job effects decrease in the first two models while the reallocation goes in the opposite direction in the second two models. The job differences, however, remain the same. The age and educational effects also undergo minor changes when the models are divided and interactions are introduced.

The interaction age x education is rather small, but the same pattern is found in all four models. Employees with academic training receive more than average for young employees and academics and less than average for old employees and academics, i.e. salary differences between academics and non-academics *decrease* by age. An example from research and development is given in figure 5:16. This interaction is perhaps strongest for economists in accounting, economic work and office work. The average difference between employees with a degree in business & economics and those with a certificate in commerce is estimated at 20 % in the age group less than 30 years but only at 7 % in the group over 44 years. This result may seem to be the opposite of what was found in chapter 4 where the differences between educational qualifications increased by age, but there is no contradiction. The findings in chapter 4 were obtained without any standardization for job differences and can be explained by educational differences in promotion. When these differences are eliminated it is very likely that the educational differences decrease by age as formal education becomes more and more obsolete.

The most important interaction is found between age and job level as is illustrated in figure 5:17 for production supervision and control and in figure 5:18 for commercial jobs. At low job level the salary differences between young and old are very much less than at high levels. For instance, an employee who has not been promoted from a low level job in production work when he

Figure 5:16. *Interaction age x education in research, experimentation and laboratory work*
(model IIIc: P64)



(model IIIb: P64)

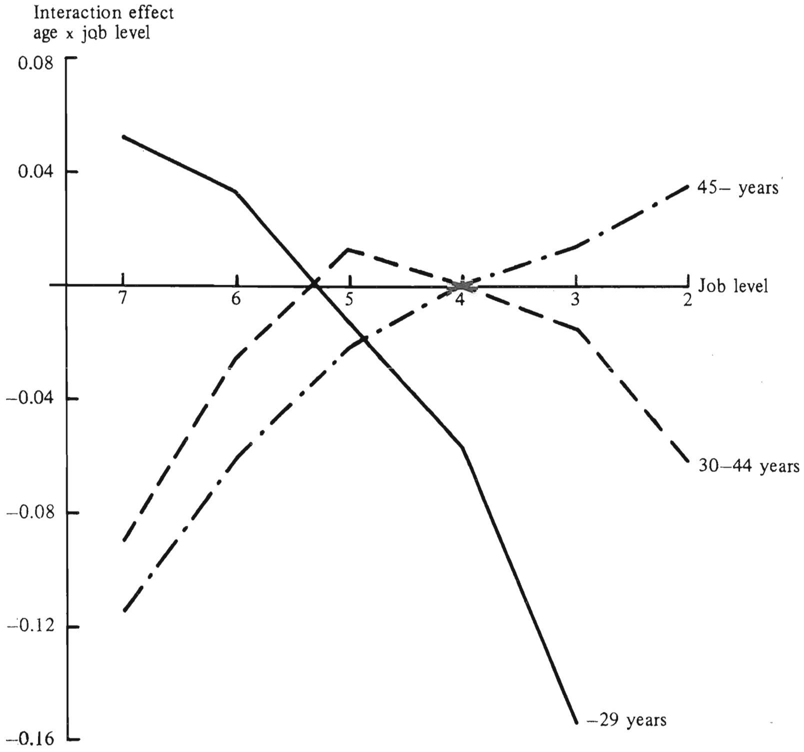
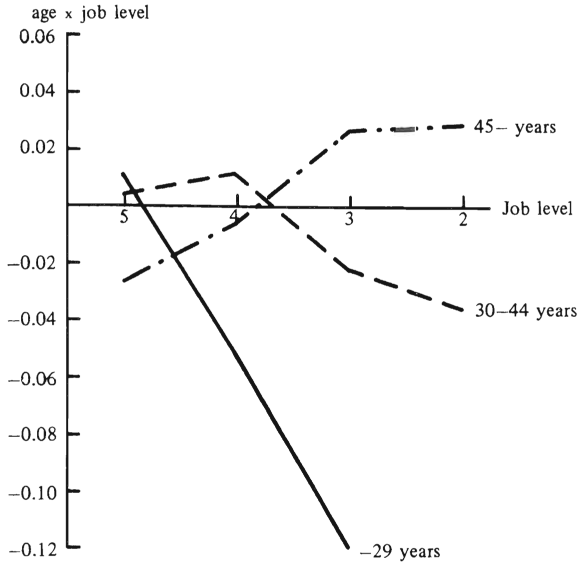


Figure 5:18. *Interaction age x job level in commercial jobs* (model VIIIe: P64,1)



is 45 or over receives on average 11 % less than is normal for his age and job level (in this example level 7), while an employee younger than 30 years receives 5 % more than normal for his age. From the figures it is also easy to see that young employees who obtain promotion to level 3 obtain very much less than normal at this job level, between 12 % and 16 % less.

The estimates of the interaction age x job family are in general very small and uncertain. The same is true for education x job level, but there it is possible to find a fairly general pattern. The salary differences between academic and non-academic educational qualifications are somewhat smaller at high job levels than at low levels, because employees without an academic degree obtain a higher salary than is normal for this group, which indicates that formal education is of less importance at top levels than at low and middle levels. An illustration is given in figures 5:19 and 5:20.

With a few exceptions education x job family mostly gives small and uncertain estimates. Figure 5:21 shows the interaction in research and development. A degree in engineering is rewarded above normal in Mechanical and electrical design and standardization (310) and in Time study and technical rationalization (400) and below normal in Mathematical and statistical work (200) and in Laboratory and experimental work (210). The opposite is true for those who posess a certificate in engineering I. Scientists receive more than normal in 200 but a little less in 210. There are no scientists in 310 and 400.

Figure 5:19. *Interaction education x job level in commercial jobs* (model VIIIe: P64,1)

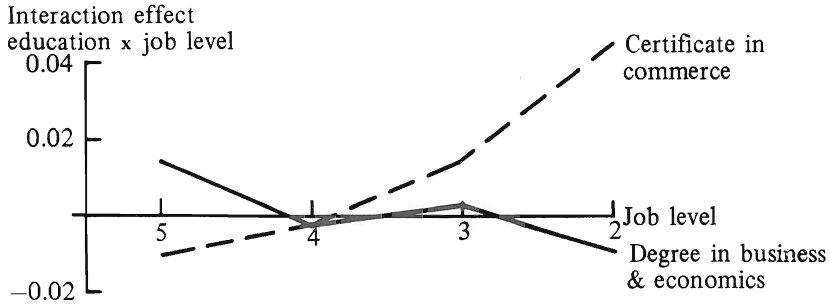


Figure 5:20. *Interaction education x job level in production supervision and control*
(model IIIb: P64)

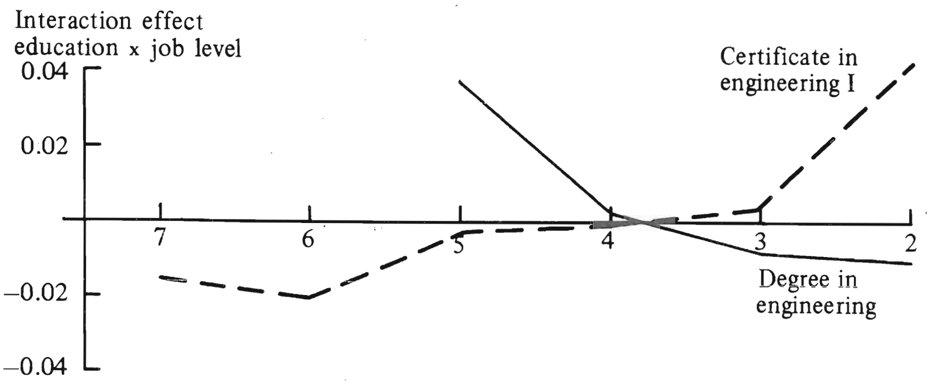
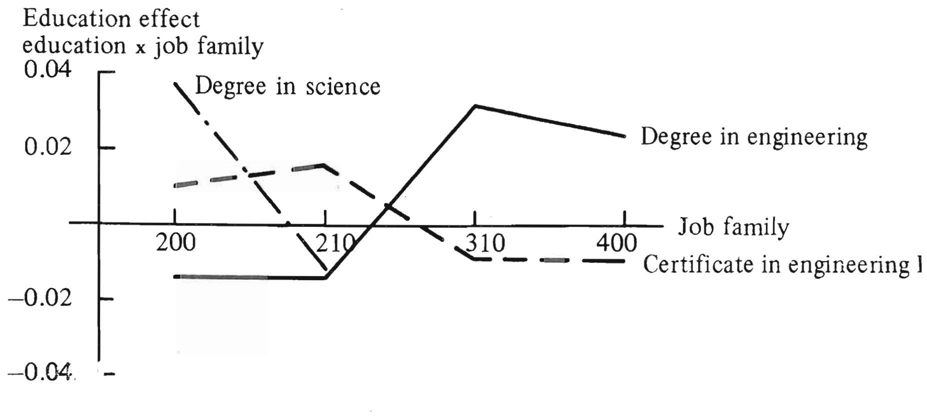


Figure 5:21. *Interaction education x job family in research experimentatio and laboratory work*
(model IIIc: P64)



5.2.6 *A comparison between Leavers, Pairs and Beginners*

The three groups, those who leave a job with an employer, those who stay with the same employer and those who take a job with a new employer, have only been identified indirectly by the method previously defined (p.42). Among those who leave a job two groups may be distinguished, those who retire and leave the labour market and those who look for another job elsewhere. The first group belongs in general to the age interval over 60 years, probably with a median age close to 65 years. With the general shape of the age-earnings profile this group should have a somewhat lower salary than the younger Pairs in the same age interval who are two–three years younger. The second group is a mix of quitters and those who are fired. It is believed that quitters form the majority during the period studied. This average salary is certainly low because quitters are usually young, but even after a standardization for age (and job) differences they may have a low salary. The explanation appears to be that they quit just because of the salary being too low. Beginners, those who take a new job, can also be divided into two groups, namely those who take their first job and those who have left another job behind. The first subgroup of course contains young employees who simply obtain an ordinary initial salary, while the second group includes a large proportion of employees who switch employers to get a better job. Is the salary they obtain above or below the average after standardization for age and job differences? One argument in favour of a relatively high salary is that this is just the reason for switching employers; one argument in favour of a relatively low salary is that promotion to a new job level (in this case combined with a switch of employer) usually does not immediately give the average salary of that level, but a salary below average. By definition Leavers and Beginners are more active (mobile) than Pairs. There are both those who are at the same time active and productive and those who are job-hoppers. The former group probably has a higher average salary than the latter.

The comparison between Leavers, Pairs and Beginners is carried out with model I which includes the three categories of engineers employed in technical jobs. This additive model is applied to data from 1963 and 1964. As the constraints used differ from group to group it is not meaningful to compare coefficient by coefficient, and contrasts are used instead. A sample of these contrasts, simple differences, is reproduced in table 5:16. A comparison between the intercepts, i.e. between the logarithmic average salaries, shows that those who leave have on average a salary 12 % lower than those who do not leave, and those who begin a job with a new employer obtain an average salary which is 27 % lower. In both cases a large proportion of young employees decreases the averages.

Table 5:16. *Comparisons between estimates of model I obtained from Leavers in 1963, Pairs in 1963–1964 and Beginners in 1964*

Comparison	Model I estimated from			
	L63	P63	P64	B64
Intercept		0.1163	0.0854	−0.2440
45–59 years and −25 years	0.3256	0.3253	0.3068	0.3725
45–59 years and 60– years	0.0694	0.0150	0.0283	0.0507
Cost of living areas 5 and 3	0.0630	0.0673	0.0686	0.0855
Degree in engineering and Certificate in engineering I	0.1092	0.1420	0.1479	0.1179
Degree in engineering and Certificate in engineering II	0.1475	0.1738	0.1773	0.1466
Production manager (1103) and Production engineer (1105)	0.4066	0.3686	0.3603	0.3889
Laboratory manager (2103) and Laboratory engineer (2105)	0.4481	0.4269	0.4240	0.4006
Senior design engineer (3103) a and Design engineer (3105)	0.4325	0.3977	0.4022	0.4069
Production engineer (1105) and Laboratory engineer (2105)	0.1088	0.1094	0.1198	0.0722
Production engineer (1105) and Design engineer (3105)	0.1062	0.1041	0.1168	0.0800
Engineering works and Mining	−0.0390	−0.0346	−0.0134	−0.0337
Ditto and Chemical industry	−0.0271	−0.0077	−0.0164	0.0004
Ditto and Building and construction	−0.0225	−0.0625	−0.0883	−0.0391
Ditto and Textile industry	0.0397	0.0049	−0.0014	−0.0022
Ditto and Leather, furs and rubber industries	0.0652	0.0322	0.0055	0.0347

The difference between the intercepts for Pairs 1963 and Pairs 1964 shows that those who stayed with their employer obtained an average salary increase of 9 %. The general average salary increase of the three groups together is influenced by retirement, new employment and mobility in general. To investigate how much mobility increases or decreases the average salary, estimates of the general averages for 1963 and 1964 would be needed. As they are not available, they are approximated by the weighted average of the intercepts for Leavers and Pairs 1963 and for Pairs and Beginners 1964, respectively. The weights are proportional to the number of observations. The result obtained is 7.8374 for 1963 and 7.8958 for 1964, i.e. the average salary has increased by approximately 6 %. With a slight simplification we may interpret the difference between 9 % and 6 % as an effect of the increased supply of young labour.

A general result from the comparisons in table 5:16 is that the differences in salary structure between the three groups are small. The results previously obtained from the frequent use of Pairs only can obviously be generalized to hold with a good approximation for all three data sets. Most of the small differences observed are difficult to comment on. The greater difference among Beginners between those who are 45–59 years and those who are less than 26 years can probably be explained by the large proportion of Beginners who take their first job, but why is the difference between, for instance, a production engineer (1105) and a design engineer (3105) less among Beginners than among Leavers and Pairs?

5.2.7 *Changes in the salary structure in the period 1957–1968*

To investigate what changes there have been in the salary structure between 1957 and 1968, model II has been estimated from the data sets »Whole 1957», »Pairs 1960 and 1964» and »Whole 1968», and model VIIIa has been estimated from the same sets except for »Whole 1957». All major job families are covered by these two models. The same constraints, No. 4, have been used to facilitate a comparison between years. Unfortunately, however, there are other differences between the data sets which do not make them perfectly comparable. The years 1957 and 1968 include Leavers, Pairs and Beginners, while 1960 and 1964 only cover Pairs. Considering the results from the previous section and the fact that Pairs are undoubtedly the most numerous group, this can only be a minor disadvantage. More serious is the change of job nomenclature in 1965, which makes it impossible to identify exactly corresponding job families before and after the change. The nature of the change has previously been commented upon in chapter 2. It is considered that this will not invalidate a comparison except for the job factor itself, although the comparison naturally has to be done with some care. Finally, there are two technical errors in the 1957 and 1960 data sets. Mechanical errors on a magnetic tape made it impossible to read approximately 900 observations in the 1957 set and approximately 300 in the 1960 set. This drop-out may explain why there are too few observations in cost of living area 5 in the first set and no observations on the job 2105 in the second set. How the lost observations are distributed across the other factors is unknown.

In addition to these errors there is the natural distribution of individual salaries which makes the estimates and then also the comparisons uncertain. This should be observed in particular when industries and jobs are compared.

From the differences in intercepts (table 5:18) it is understood that the average salary★ in technical job families increased by approximately 100 % from 1957 to 1968. Between 1960 and 1968 the increase was approximately 70 %. In the commercial, economic and administrative job families the average increase was almost as much.

To obtain a general view of the structural changes over the period 1957–1968 it is natural to compare factor ranges (table 5:17). The salary differences due to the age decreased until the mid-sixties but have increased again since then. The differentiation was about the same in 1968 as over the period 1957/1960. In the technical job families the differences between the cost of living areas have increased by a few percentage units while this is not so in the commercial, economic and administrative families. The maximum difference between educational groups has remained almost constant in both groups of families. No major changes in differentiation by job have been found either, although the conclusion

Table 5:17. *Factor ranges from model II and model VIIIa in 1957–1968*

Factor	Model II				Model VIIIa		
	H57	P60	P64	H68	P60	P64	H68
Age	0.3728	0.3513	0.3088	0.3606	0.4018	0.3769	0.4057
Cost of living area	0.0597	0.0823	0.0674	0.1039	0.0700	0.0534	0.0650
Education	0.1618	0.1755	0.1749	0.1676	0.1713	0.1699	0.1634
Job	0.9984	0.9977	0.8850	(0.9824)	1.0365	1.0250	–
Industry	0.0997	0.1016	0.0993	0.0492	0.1270	0.0947	0.0552

★ See section 5.1.2.2 for the definition of average salary when constraints No. 4 are used.

is uncertain because of insufficient comparability. Both models, however, indicate a decrease in differences between industries of approximately 50 %.

The general conclusion from this rough survey is that no major changes in the salary structure have taken place. To qualify this conclusion it is, however, desirable to go into more details. To this end table 5:18 has been calculated where a comparison is made estimate by estimate.

The decrease in differentiation between young and old employees obtained in 1964 almost vanished by 1968. A comparison between the age-effect estimates in tables II:P64,3 and II:H68,3 in appendix A indicates that the increased inequality in 1968 is at least partly due to a decrease of salaries for the youngest rather than an increase for the old and middle aged employees which in turn fits very well with the increased supply of young labour from schools and universities (see chapter 3) and what we know of the general labour market conditions for this type of labour during the end of the sixties.

Although the changes are small, salaries in the technical job families increased more in cost of living area 5 than in areas 4 and 3. The change occurred during the first and last part of the period.

Between 1957 and 1960 employees with a degree in engineering improved their relative salary position while in particular those with a degree in science decreased theirs. The same situation prevailed between 1960 and 1964. In 1968 the situation is changed. The excess supply of engineers has resulted in a decrease in the relative salary position of engineers, in particular compared to employees with »other» university degrees. For the whole period the result is that the salaries of employees with a degree in science decreased from 5 % to 1 % above average, while employees with »other» university degrees experienced an increase from 1 % to 5 % above average. The relative positions of the engineers were almost the same in 1957 as in 1968. All these results are valid for the technical job families. The estimates of model VIIIa reveal only very small changes.

The decreased differences between industries observed above hide some major changes in relative positions. Employees in technical jobs in Mining obtained 6 % more than average in 1957 but just the average in 1968. Those in commercial, economic and administrative jobs experienced a similar decrease from 8 % above average in 1960 to 1 % above average in 1968. Technical jobs in Building and construction were paid 2 % less than average in 1957, 5 % more than average in 1960, 7 % more than average in 1964 and only 2 % more than average in 1968. Commercial, economic and administrative jobs were paid 3 % less than average in 1960 and 2 % more than average in 1968. In Printing and allied industries technical jobs were paid about average in 1957 but 4 % below average in 1968, while commercial, economic and administrative jobs were paid 5 % less than average in 1960 and 3 % above average in 1968. Other changes

Table 5:18. *Differences between years in estimates of model II and model VIIIa*

	Model II				Model VIIIa		
	P60–H57	P64–P60	H68–P64	H68–H57	P64–P60	H68–P64	H68–P60
Intercept	0.185	0.289	0.232	0.706	0.301	0.253	0.554
Age							
–25	0.014	0.025	–0.037	0.002	0.019	–0.025	–0.006
26–29	–0.005	0.024	–0.003	0.016	0.010	–0.008	0.002
30–34	0.006	–0.004	0.006	0.008	0.003	–0.003	0.000
35–44	0.004	–0.006	0.009	0.007	0.005	0.003	0.008
45–59	–0.007	–0.018	0.015	–0.010	–0.006	0.004	–0.002
60–	–0.012	–0.021	0.009	–0.024	–0.032	0.029	–0.003
Cost of living area							
3	–0.008	0.010	–0.019	–0.017	0.012	–0.011	0.001
4	–0.007	–0.005	0.001	–0.011	–0.006	0.010	0.004
5	0.015	–0.005	0.018	0.028	–0.005	0.001	–0.004
Education							
Degree in engineering	0.015	0.006	–0.014	0.007	–0.005	0.001	–0.004
Certificate in engineering I	0.011	–0.008	–0.011	–0.008	–0.004	–0.002	–0.006
Certificate in engineering II	0.001	0.007	–0.007	0.001	0.010	–0.001	0.009
Degree in business & economics					0.005	–0.006	–0.001
Certificate in commerce					–0.005	0.008	0.003
Degree in science	–0.030	–0.010	–0.001	–0.041			
Other academic degrees	0.003	0.004	0.033	0.040			
Industry							
Mining	–0.038	–0.021	–0.006	–0.065	–0.024	–0.040	–0.064
Metal and engineering industry							
Iron and steel works, metal plants	0.018	–0.016	–0.017	–0.015	–0.039	–0.015	–0.054
Manufacture of hardware	–0.018	–0.010	–0.014	–0.014	0.014	–0.003	0.011
Engineering works	0.014	–0.003	–0.006	0.005	–0.001	–0.009	–0.010
Repair works	–0.014	–0.005	–0.004	–0.023	–0.005	0.021	0.016
Shipyards	–	–0.025	–0.006	–	–0.047	–0.012	–0.059
Manufacture of electrical equipment	–0.031	0.025	–0.002	–0.008	0.010	0.013	0.023
Other metal industry	–0.010	0.030	0.008	0.028	–0.010	0.002	–0.008
Quarrying: stone, clay and glass products	0.021	–0.002	–0.002	0.017	0.020	–0.022	–0.002
Wood industry	0.008	0.004	0.000	0.012	0.004	0.038	0.042
Manufacture of pulp, paper and paper products	0.049	–0.026	–0.002	0.021	–0.019	0.004	–0.015
Printing and allied industries	–0.020	–0.011	–0.011	–0.042	0.013	0.063	0.076
Food manufacturing industries	0.012	–0.014	–0.002	–0.004	0.008	0.025	0.033
Beverage and tobacco industries	–0.047	0.038	0.026	0.017	0.037	–0.028	0.009
Textile industry	–0.026	0.008	0.002	–0.016	–0.013	0.002	–0.011
Leather, furs and rubber industries	–0.029	0.004	0.006	–0.019	0.002	–0.016	–0.014
Chemical industry	0.032	0.000	–0.002	0.030	–0.005	–0.014	–0.019
Building and construction	0.065	0.026	–0.049	0.042	0.055	–0.009	0.046

to notice are a decrease in the relative position for Shipyards; Iron and steel works and metal plants and an increase for Wood industry and the Food manufacturing industries.

The comparisons for the job factor have to be limited to the period 1957–1964. This is done in Klevmarken [1968a]. Increases and decreases are scattered across the job families and the job levels without any very clear pattern. A few changes may, however, be noted. In production supervision and control (110) the differences between the job levels decreased, and a similar decrease occured in the commercial, economic and administrative jobs except in 810.

Although the more detailed analysis of changes in the salary structure has revealed that changes do occur, the general impression of a fairly permanent salary structure still remains. Is this result remarkable or is it only what might be expected? During the sample period changing profit margins in Swedish industry have caused a reallocation of labour which is partly reflected in the changes of composition described in chapter 3. One example is the closing down of a large part of the Swedish textile industry and the increased activity in Building and construction. For this reason one may be inclined to believe that these structural changes would also be reflected in a new salary structure.

There are two important observations to make before one proceeds to a discussion about the causes underlying the present salary structure. Firstly, the salary structure is measured by the average salaries of all employees, not only those who change their job or employer. As the dominant majority of the employees keep the same employment for more than one or two years the average salaries cannot but change slowly. Secondly, the salary structure is now defined by a number of factors such as job and industry. A reallocation of employees between jobs and industries will not in principle change the salary structure unless the average salaries in some factor combinations also change.

One possible explanation of the observed permanence in salary structure is that there are market forces working towards this structure. For instance, the well trained group of employees included in this study may because of its mobility have been able to get compensation for relative salary decreases simply by moving to branches and jobs with higher salaries. This adjustment mechanism also implies that the average salary increases in an industry or a job from which labour moves, because those who stay already have a good salary and thus have no incentive to move. If this hypothesis is correct, it would be possible to observe more changes in salary structure for another less mobile group of salary earners.

Another explanation is that the salary structure is rigid because the bargaining system contains so much institutional restrictions that the salary structure is maintained almost unchanged. The importance of salary statistics in this system and as a means whereby an individual employer (and employee) adjusts to the present structure has already been stressed.

It is an interesting task for future research to investigate the relative importance of these two alternative explanations for the observed permanence in the salary structure. An attempt to explain the lack of changes also implies suggestions of potential changes which closely agrees with the extended profile analysis suggested in chapter 4, section 4.3.7.

APPENDIX A
MODELS AND ESTIMATES

For each model and application the number of observations are first tabled by factors and interactions. Then follow the ordinary least squares estimates and inside brackets their respective standard errors.

For a survey of the models and an explanation of the notation used see section 5.2. Definitions of variables and a descriptive account of the data have been given in chapter 2.

Contents

TABLE IIa:H57, number of observations

Total number of observations: 15,525

Age	
– 25	1,167
26 – 29	2,336
30 – 34	3,649
35 – 44	5,480
45 – 59	2,445
60 –	488

Cost of living area	
2 and 3	11,826
4	3,416
5	283

Education	
Degree in engineering	2,470
Certificate in engineering I	5,727
Certificate in engineering II	7,153
Degree in science	117
Other degrees	58

Job

	Level					
Family	2	3	4	5	6	7
Supervision of production	314	666	1,184	929		
Work supervision, general			194	343		
Work supervision, building and construction work				158	253	123
Mathematical work	7	79	232	228	65	
Laboratory work	110	370	582	1,006	308	21
Mechanical and electrical design engineering	150	484	1,194	2,067	1,444	345
Productivity engineering	30	129	267	578	144	
Technical instruction and technical service		44	151	207	29	
Other technical work	60	204	451	375		

Industry	
Mining	362
Metal and engineering industry	
Iron and steel works, metal plants	1,333
Manufacture of hardware	576
Engineering works	5,208
Repair works	56
Shipyards	1,031
Manufacture of electrical equipment	1,718
Other metal industry	170
Quarrying: stone, clay and glass products	426
Wood industry	171
Manufacture of pulp, paper and paper products	970
Printing and allied industries	0
Food manufacturing industries	271
Beverage and tobacco industries	65
Textile industry	552
Leather, furs and rubber industries	191
Chemical industry	1,001
Building and construction	1,424

Note. Mechanical errors on the magnetic tape made it impossible to read 900 observations.

TABLE IIa:H57,3

Intercept 7.5388 (0.0071)

R = 0.8950

Age		
– 25	−0.2223	(0.0047)
26 – 29	−0.1220	(0.0034)
30 – 34	−0.0175	(0.0028)
35 – 44	0.0814	(0.0026)
45 – 59	0.1505	(0.0034)
60 –	0.1300	(0.0065)

Cost of living area		
3	−0.0286	(0.0038)
4	−0.0024	(0.0040)
5	0.0311	(0.0072)

Education		
Degree in engineering	0.0761	(0.0060)
Certificate in engineering I	−0.0517	(0.0059)
Certificate in engineering II	−0.0857	(0.0060)
Degree in science	0.0507	(0.0130)
Other degrees	0.0106	(0.0174)

Job

Family	Level 2	3	4	5	6	7
Supervision of production	0.5422 (0.0095)	0.3247 (0.0068)	0.1145 (0.0054)	−0.0864 (0.0059)		
Work supervision, general			−0.0106 (0.0120)	−0.1321 (0.0096)		
Work supervision, building and construction work				−0.2091 (0.0128)	−0.2656 (0.0109)	−0.3765 (0.0149
Mathematical work	0.4851 (0.0587)	0.2621 (0.0179)	0.0109 (0.0108)	−0.1866 (0.0107)	−0.3091 (0.0196)	
Laboratory work	0.4402 (0.0155)	0.2166 (0.0091)	−0.0109 (0.0072)	−0.2245 (0.0058)	−0.3244 (0.0097)	−0.4562 (0.0340
Mechanical and electrical design engineering	0.4519 (0.0132)	0.1979 (0.0077)	−0.0238 (0.0054)	−0.1904 (0.0045)	−0.3152 (0.0053)	−0.4478 (0.0092
Productivity engineering	0.3732 (0.0284)	0.1453 (0.0140)	−0.0519 (0.0099)	−0.2007 (0.0072)	−0.3309 (0.0133)	
Technical instruction and technical service		0.3234 (0.0236)	0.0208 (0.0130)	−0.1832 (0.0112)	−0.2855 (0.0290)	
Other technical work	0.5348 (0.0204)	0.2748 (0.0113)	0.0568 (0.0079)	−0.1538 (0.0086)		

Industry

Mining	0.0644	(0.0092)
Metal and engineering industry		
Iron and steel works metal plants	0.0261	(0.0049)
Manufacture of hardware	0.0195	(0.0077)
Engineering works	−0.0258	(0.0117)
Repair works	0.0317	(0.0055)
Shipyards	–	–
Manufacture of electrical equipment	−0.0039	(0.0095)
Other metal industry	−0.0353	(0.0188)
Quarrying: stone, clay and glass products	−0.0011	(0.0069)
Wood industry	−0.0103	(0.0111)
Manufacture of pulp, paper and paper products	−0.0118	(0.0055)
Printing and allied industries	0.0058	(0.0055)
Food manufacturing industries	−0.0209	(0.0068)
Beverage and tobacco industries	−0.0068	(0.0034)
Textile industry	0.0048	(0.0204)
Leather furs and rubber industries	0.0094	(0.0056)
Chemical industry	−0.0290	(0.0046)
Building and construction	−0.0170	(0.0117)

	Sum of squares	d.f.
Total	1960.1267	15,524
Explained	1570.0993	65
Residual	390.0274	15,459

TABLE IIa: P60, number of observations

Total number of observations: 13,824

Age

– 25	723
26 – 29	1,365
30 – 34	2,373
35 – 44	5,789
45 – 59	3,101
60 –	473

Cost of living area

3	7,942
4	2,529
5	3,353

Education

Degree in engineering	2,771
Certificate in engineering I	4,467
Certificate in engineering II	6,343
Degree in science	134
Other degrees	109

Job

Family	Level 2	3	4	5	6	7
Supervision of production	300	717	1,326	1,116		
Work supervision, general			222	530		
Work supervision, building and construction work				107	419	208
Mathematical work	15	96	275	14	9	
Laboratory work	128	533	898	0	511	8
Mechanical and electrical design engineering	146	602	1,646	1,023	297	343
Productivity engineering	42	140	355	20	2	
Technical instruction and technical service		66	229	337	0	
Other technical work	52	218	494	380		

Industry

Mining	284
Metal and engineering industry	
Iron and steel works, metal plants	1,057
Manufacture of hardware	376
Engineering works	3,898
Repair works	72
Shipyards	609
Manufacture of electrical equipment	2,209
Other metal industry	140
Quarrying: stone, clay and glass products	377
Wood industry	133
Manufacture of pulp, paper and paper products	764
Printing and allied industries	33
Food manufacturing industries	221
Beverage and tobacco industries	93
Textile industry	303
Leather, furs and rubber industries	149
Chemical industry	875
Building and construction	2,231

TABLE IIa:P60,1

Intercept 7.7152 (0.0013)

R = 0.8951

Age

− 25	−0.2548	(0.0066)
26 − 29	−0.1734	(0.0042)
30 − 34	−0.0579	(0.0029)
35 − 44	0.0390	(0.0016)
45 − 59	0.0965	(0.0025)
60 −	0.0710	(0.0069)

Cost of living area

3	−0.0247	(0.0012)
4	0.0013	(0.0028)
5	0.0576	(0.0025)

Education

Degree in engineering	0.1242	(0.0032)
Certificate in engineering I	−0.0069	(0.0019)
Certificate in engineering II	−0.0513	(0.0016)
Degree in science	0.0540	(0.0014)
Other degrees	0.0468	(0.0155)

Job

Family	Level					
	2	3	4	5	6	7
Supervision of production	0.5607	0.3125	0.1068	−0.0758		
	(0.0089)	(0.0056)	(0.0040)	(0.0044)		
Work supervision, general			−0.0053	−0.1425		
			(0.0106)	(0.0072)		
Work supervision, building and construction work				−0.2339	−0.2799	−0.4267
				(0.0146)	(0.0082)	(0.0116
Mathematical work	0.5197	0.2512	0.0421	−0.1653	−0.3492	
	(0.0389)	(0.0156)	(0.0094)	(0.0403)	(0.0503)	
Laboratory work	0.4502	0.2254	0.0184	−	−0.2877	−0.4358
	(0.0137)	(0.0071)	(0.0051)		(0.0073)	(0.0532
Mechanical and electrical design engineering	0.4602	0.2066	−0.0169	−0.1828	−0.3307	−0.4060
	(0.0126)	(0.0062)	(0.0037)	(0.0047)	(0.0090)	(0.0089
Productivity engineering	0.3625	0.1555	−0.0417	−0.1181	−0.3750	
	(0.0232)	(0.0127)	(0.0079)	(0.0338)	(0.0707)	
Technical instruction and technical service		0.2730	0.0350	−0.1830	−	
		(0.0186)	(0.0102)	(0.0082)		
Other technical work	0.4818	0.2698	0.0488	−0.1422		
	(0.0210)	(0.0103)	(0.0068)	(0.0077)		

Industry

Mining	0.0219	(0.0089)
Metal and engineering industry		
Iron and steel works, metal plants	0.0395	(0.0046)
Manufacture of hardware	−0.0030	(0.0077)
Engineering works	−0.0160	(0.0022)
Repair works	0.0132	(0.0178)
Shipyards	0.0092	(0.0062)
Manufacture of electrical equipment	−0.0399	(0.0032)
Other metal industry	−0.0499	(0.0127)
Quarrying: stone, clay and glass products	0.0152	(0.0077)
Wood industry	−0.0064	(0.0130)
Manufacture of pulp, paper and paper products	0.0327	(0.0055)
Printing and allied industries	−0.0189	(0.0262)
Food manufacturing industries	−0.0139	(0.0102)
Beverage and tobacco industries	−0.0583	(0.0156)
Textile industry	−0.0257	(0.0087)
Leather, furs and rubber industries	−0.0242	(0.0123)
Chemical industry	−0.0019	(0.0053)
Building and construction	0.0432	(0.0041)

	Sum of squares	d.f.
Total	1555.3711	13,823
Explained	1246.0488	63
Residual	309.3224	13,760

TABLE IIa:P60,3

Intercept 7.7240 (0.0061)

R = 0.8951

Age		
– 25	–0.2082	(0.0058)
26 – 29	–0.1268	(0.0040)
30 – 34	–0.0113	(0.0032)
35 – 44	0.0856	(0.0027)
45 – 59	0.1431	(0.0032)
60 –	0.1176	(0.0061)

Cost of living area		
3	–0.0361	(0.0019)
4	–0.0100	(0.0024)
5	0.0462	(0.0023)

Education		
Degree in engineering	0.0908	(0.0049)
Certificate in engineering I	–0.0403	(0.0048)
Certificate in engineering II	–0.0847	(0.0049)
Degree in science	0.0207	(0.0112)
Other degrees	0.0135	(0.0126)

Job

Family	Level 2	3	4	5	6	
Supervision of production	0.5453 (0.0095)	0.2972 (0.0068)	0.0915 (0.0056)	–0.0911 (0.0058)		
Work supervision, general			–0.0206 (0.0109)	–0.1579 (0.0080)		
Work supervision, building and construction work				–0.2494 (0.0147)	–0.2952 (0.0088)	–0. (0.
Mathematical work	0.5036 (0.0380)	0.2358 (0.0157)	0.0268 (0.0099)	–0.1814 (0.0393)	–0.3657 (0.0490)	
Laboratory work	0.4348 (0.0139)	0.2100 (0.0079)	0.0030 (0.0064)	–	–0.3031 (0.0081)	–0. (0.
Mechanical and electrical design engineering	0.4448 (0.0129)	0.1913 (0.0072)	–0.0322 (0.0054)	–0.1918 (0.0061)	–0.3461 (0.0096)	–0. (0.
Productivity engineering	0.3469 (0.0229)	0.1402 (0.0130)	–0.0570 (0.0087)	–0.1340 (0.0331)	–0.3887 (0.0700)	
Technical instruction and technical service		0.2575 (0.0185)	0.0196 (0.0107)	–0.1983 (0.0089)	–	
Other technical work	0.4663 (0.0208)	0.2544 (0.0108)	0.0335 (0.0078)	–0.1575 (0.0085)		

Industry

Mining	0.0265	(0.0089)
Metal and engineering industry		
Iron and steel works, metal plants	0.0441	(0.0052)
Manufacture of hardware	0.0016	(0.0078)
Engineering works	–0.0114	(0.0036)
Repair works	0.0178	(0.0171)
Shipyards	0.0138	(0.0066)
Manufacture of electrical equipment	–0.0353	(0.0043)
Other metal industry	–0.0453	(0.0123)
Quarrying: stone, clay and glass products	0.0198	(0.0078)
Wood industry	–0.0018	(0.0125)
Manufacture of pulp, paper and paper products	0.0373	(0.0059)
Printing and allied industries	–0.0143	(0.0249)
Food manufacturing industries	–0.0093	(0.0100)
Beverage and tobacco industries	–0.0537	(0.0150)
Textile industry	–0.0210	(0.0086)
Leather, furs and rubber industries	–0.0196	(0.0120)
Chemical industry	0.0027	(0.0058)
Building and construction	0.0479	(0.0050)

	Sum of squares	d.f.
Total	1555.3711	13,823
Explained	1246.0616	63
Residual	309.3095	13,760

TABLE IIa: P64, number of observations

Age

– 25	1,314
26 – 29	3,159
30 – 34	4,100
35 – 44	8,621
45 – 59	5,753
60 –	707

Cost of living area

3	14,152
4	4,199
5	5,303

Education

Degree in engineering	3,617
Certificate in engineering I	8,136
Certificate in engineering II	11,530
Degree in science	207
Other degrees	164

Job

Family	Level 2	3	4	5	6	7
Supervision of production	413	1,052	1,603	1,208		
Work supervision, general			255	582		
Work supervision, building and construction work				166	456	71
Mathematical work	16	106	372	341	148	
Laboratory work	140	601	1,305	1,947	591	93
Mechanical and electrical design engineering	138	744	2,087	3,156	1,722	309
Productivity engineering	40	190	464	746	242	
Technical instruction and technical service		81	329	472	85	
Other technical work	56	267	607	453		

Industry

Mining	395
Metal and engineering industry	
Iron and steel works, metal plants	1,759
Manufacture of hardware	706
Engineering works	7,384
Repair works	116
Shipyards	1,057
Manufacture of electrical equipment	4,424
Other metal industry	209
Quarrying: stone, clay and glass products	653
Wood industry	223
Manufacture of pulp, paper and paper products	1,251
Printing and allied industries	90
Food manufacturing industries	350
Beverage and tobacco industries	139
Textile industry	409
Leather, furs and rubber industries	273
Chemical industry	1,498
Building and construction	2,718

TABLE IIa:P64,1

Intercept 7.9444 (0.0008)

R = 0.9121

Age

– 25	–0.2192	(0.0039)
26 – 29	–0.1392	(0.0024)
30 – 34	–0.0512	(0.0019)
35 – 44	0.0439	(0.0012)
45 – 59	0.0896	(0.0016)
60 –	0.0607	(0.0049)

Cost of living area

3	–0.0172	(0.0007)
4	–0.0054	(0.0019)
5	0.0502	(0.0017)

Education

Degree in engineering	0.1368	(0.0024
Certificate in engineering I	–0.0091	(0.0012
Certificate in engineering II	–0.0382	(0.001
Degree in science	0.0508	(0.009
Other degrees	0.0575	(0.010

Job

Family	Level 2	3	4	5	6	
Supervision of production	0.5507	0.3495	0.1571	–0.0109		
	(0.0066)	(0.0041)	(0.0032)	(0.0037)		
Work supervision, general			0.0563	–0.0968		
			(0.0085)	(0.0059)		
Work supervision, building and construction work				–0.1638	–0.2584	–
				(0.0102)	(0.0067)	(
Mathematical work	0.4969	0.3320	0.1012	–0.1016	–0.2768	
	(0.0327)	(0.0128)	(0.0070)	(0.0071)	(0.0109)	
Laboratory work	0.5047	0.2930	0.0824	–0.1299	–0.2528	–
	(0.0113)	(0.0057)	(0.0037)	(0.0029)	(0.0056)	(
Mechanical and electrical design engineering	0.5154	0.2752	0.0548	–0.1273	–0.2761	–
	(0.0113)	(0.0048)	(0.0028)	(0.0022)	(0.0033)	(
Productivity engineering	0.4331	0.2079	0.0319	–0.1380	–0.2606	
	(0.0206)	(0.0094)	(0.0060)	(0.0048)	(0.0084)	
Technical instruction and technical service		0.3533	0.0870	–0.1070	–0.2523	
		(0.0146)	(0.0073)	(0.0060)	(0.0142)	
Other technical work	0.5352	0.3140	0.1090	–0.0704		
	(0.0175)	(0.0080)	(0.0053)	(0.0061)		

Industry

Mining	0.0020	(0.0066)
Metal and engineering industry		
Iron and steel works, metal plants	–0.0245	(0.0031)
Manufacture of hardware	–0.0125	(0.0049)
Engineering works	–0.0183	(0.0014)
Repair works	0.0089	(0.0122)
Shipyards	–0.0151	(0.0041)
Manufacture of electrical equipment	–0.0136	(0.0019)
Other metal industry	–0.0195	(0.0090)
Quarrying: stone, clay and glass products	0.0141	(0.0051)
Wood industry	–0.0017	(0.0087)
Manufacture of pulp, paper and paper products	0.0071	(0.0037)
Printing and allied industries	–0.0294	(0.0138)
Food manufacturing industries	–0.0268	(0.0070)
Beverage and tobacco industries	–0.0191	(0.0111)
Textile industry	–0.0168	(0.0065)
Leather, furs and rubber industries	–0.0194	(0.0079)
Chemical industry	–0.0011	(0.0035)
Building and construction	0.0699	(0.0031)

	Sum of squares	d.f.
Total	2379.2523	23,653
Explained	1979.2208	71
Residual	400.0315	23,582

TABLE IIa:P64,3

Intercept 8.0133 (0.0036)

R = 0.9119

Age		
– 25	−0.1833	(0.0035)
26 – 29	−0.1033	(0.0024)
30 – 34	−0.0152	(0.0021)
35 – 44	0.0798	(0.0018)
45 – 59	0.1255	(0.0020)
60 –	0.0966	(0.0043)

Cost of living area		
3	−0.0264	(0.0012)
4	−0.0146	(0.0016)
5	0.0410	(0.0016)

Education		
Degree in engineering	0.0972	(0.0035)
Certificate in engineering I	−0.0487	(0.0033)
Certificate in engineering II	−0.0777	(0.0034)
Degree in science	0.0112	(0.0078)
Other degrees	0.0180	(0.0087)

Job

Family	Level					
	2	3	4	5	6	7
Supervision of production	0.4909	0.2897	0.0973	−0.0707		
	(0.0067)	(0.0044)	(0.0037)	(0.0041)		
Work supervision, general			−0.0034	−0.1566		
			(0.0085)	(0.0061)		
Work supervision, building and construction work				−0.2236	−0.3182	−0.429
				(0.0101)	(0.0068)	(0.015
Mathematical work	0.4372	0.2722	0.0414	−0.1614	−0.3365	
	(0.0319)	(0.0126)	(0.0070)	(0.0071)	(0.0107)	
Laboratory work	0.4450	0.2333	0.0227	−0.1897	−0.3126	−0.343
	(0.0111)	(0.0059)	(0.0040)	(0.0035)	(0.0058)	(0.013
Mechanical and electrical design engineering	0.4557	0.2155	−0.0050	−0.1871	−0.3358	−0.394
	(0.0111)	(0.0050)	(0.0033)	(0.0029)	(0.0038)	(0.007
Productivity engineering	0.3733	0.1481	−0.1278	−0.1978	−0.3203	
	(0.0202)	(0.0094)	(0.0062)	(0.0050)	(0.0085)	
Technical instruction and technical service		0.2936	0.0272	−0.1668	−0.3120	
		(0.0143)	(0.0074)	(0.0062)	(0.0140)	
Other technical work	0.4754	0.2542	0.0492	−0.1302		
	(0.0172)	(0.0080)	(0.0055)	(0.0063)		

Industry

Mining	0.0057	(0.0065)
Metal and engineering industry		
Iron and steel works metal plants	0.0282	(0.0034)
Manufacture of hardware	−0.0088	(0.0050)
Engineering works	−0.0147	(0.0023)
Repair works	0.0126	(0.0116)
Shipyards	−0.0114	(0.0043)
Manufacture of electrical equipment	−0.0099	(0.0027)
Other metal industry	−0.0158	(0.0087)
Quarrying: stone, clay and glass products	0.0178	(0.0051)
Wood industry	0.0020	(0.0084)
Manufacture of pulp, paper and paper products	0.0108	(0.0039)
Printing and allied industries	−0.0257	(0.0131)
Food manufacturing industries	−0.0231	(0.0068)
Beverage and tobacco industries	−0.0154	(0.0107)
Textile industry	−0.0131	(0.0064)
Leather furs and rubber industries	−0.0157	(0.0077)
Chemical industry	0.0026	(0.0038)
Building and construction	−0.0736	(0.0035)

	Sum of squares	d.f.
Total	2379.2523	23,653
Explained	1978.5735	66
Residual	400.6788	23,587

TABLE IIa:H68, number of observations

Total number of observations: 34,540

Age

– 25	3,586
26 – 29	5,511
30 – 34	6,030
35 – 44	9,798
45 – 59	8,784
60 –	831

Cost of living area

3	20,623
4	5,837
5	8,080

Education

Degree in engineering	4,660
Certificate in engineering I	12,220
Certificate in engineering II	17,117
Degree in science	287
Other degrees	256

Job

Family	Level					
	2	3	4	5	6	7
Administration of plants or branches	73	94	72	14		
General analytical and planning work	25	80	152	77		
Supervision of production	603	1,397	2,286	1,923	324	
Work supervision, general		30	62	368	216	25
Work supervision, building and construction work		41	300	789	967	354
Mathematical work	17	115	394	434	223	
Laboratory work	163	724	1,566	2,303	959	120
Mechanical and electrical design engineering	155	784	2,582	4,034	2,746	609
Construction and civil engineering design	25	173	441	726		
Productivity engineering	76	330	740	1,326	786	
Technical instruction and technical service	12	114	431	821	293	15

Industry

Mining	1,067
Metal and engineering industry	
Iron and steel works, metal plants	2,867
Manufacture of hardware	953
Engineering works	10,063
Repair works	130
Shipyards	1,231
Manufacture of electrical equipment	6,228
Other metal industry	339
Quarrying: stone, clay and glass products	1,009
Wood industry	396
Manufacture of pulp, paper and paper products	1,307
Printing and allied industries	100
Food manufacturing industries	535
Beverage and tobacco industries	68
Textile industry	398
Leather, furs and rubber industries	343
Chemical industry	2,235
Building and construction	5,271

TABLE IIa: H68,3

Intercept 8.2449 (0.0033)

R = 0.9319

Age

– 25	−0.2201	(0.0024)
26 – 29	−0.1061	(0.0018)
30 – 34	−0.0092	(0.0017)
35 – 44	0.0890	(0.0015)
45 – 59	0.1405	(0.0016)
60 –	0.1059	(0.0038)

Cost of living area

3	−0.0450	(0.0010)
4	−0.0140	(0.0013)
5	0.0589	(0.0012)

Education

Degree in engineering	0.0832	(0.0028)
Certificate in engineering I	−0.0596	(0.0027)
Certificate in engineering II	−0.0844	(0.0027)
Degree in science	0.0107	(0.0064)
Other degrees	0.0501	(0.0068)

Job

Family	Level 2	3	4	5	6	7	8
Administration of plants or branches	0.6710 (0.0149)	0.4421 (0.0131)	0.1766 (0.0149)	−0.1106 (0.0335)			
General analytical and planning work	0.4558 (0.0251)	0.3261 (0.0142)	0.0940 (0.0104)	−0.1164 (0.0144)			
Supervision of production	0.5094 (0.0056)	0.3005 (0.0040)	0.1093 (0.0034)	−0.0616 (0.0035)	−0.2621 (0.0073)		
Work supervision, general		0.1879 (0.0229)	−0.0043 (0.0160)	−0.2026 (0.0069)	−0.3411 (0.0089)	−0.4200 (0.0251)	
Work supervision, building and construction work		0.1251 (0.0197)	0.0157 (0.0078)	−0.1575 (0.0054)	−0.3332 (0.0051)	−0.4730 (0.0075)	
Mathematical work	0.4633 (0.0304)	0.2629 (0.0120)	0.0451 (0.0067)	−0.1497 (0.0064)	−0.3955 (0.0088)		
Laboratory work	0.4725 (0.0102)	0.2639 (0.0054)	0.0399 (0.0039)	−0.1676 (0.0034)	−0.3330 (0.0048)	−0.4385 (0.0117)	−0.7047 (0.0396)
Mechanical and electrical design engineering	0.4651 (0.0104)	0.2542 (0.0050)	0.0088 (0.0033)	−0.1856 (0.0029)	−0.3280 (0.0034)	−0.4410 (0.0057)	−0.4974 (0.0274)
Construction and civil engineering design	0.4935 (0.0251)	0.2578 (0.0098)	0.0211 (0.0063)	−0.1740 (0.0051)			
Productivity engineering	0.3927 (0.0145)	0.1640 (0.0072)	−0.0119 (0.0051)	−0.1836 (0.0041)	−0.3215 (0.0051)		
Technical instruction and technical service	0.4112 (0.0361)	0.2706 (0.0119)	0.0263 (0.0064)	−0.1530 (0.0049)	−0.3175 (0.0077)	−0.4416 (0.0324)	

Industry

Mining	−0.0003	(0.0041)
Metal and engineering industry		
Iron and steel works, metal plants	0.0115	(0.0028)
Manufacture of hardware	0.0054	(0.0042)
Engineering works	−0.0208	(0.0021)
Repair works	0.0081	(0.0107)
Shipyards	−0.0172	(0.0039)
Manufacture of electrical equipment	−0.0120	(0.0024)
Other metal industry	−0.0076	(0.0067)
Quarrying: stone, clay and glass products	0.0157	(0.0041)
Wood industry	0.0023	(0.0063)
Manufacture of pulp, paper and paper products	0.0090	(0.0037)
Printing and allied industries	0.0150	(0.0150)
Food manufacturing industries	−0.0248	(0.0055)
Beverage and tobacco industries	0.0111	(0.0121)
Textile industry	−0.0107	(0.0063)
Leather, furs and rubber industries	−0.0092	(0.0067)
Chemical industry	0.0003	(0.0031)
Building and construction	0.0244	(0.0029)

TABLE IIb:H68,3

Intercept 8.2399 (0.0044)

R = 0.8135

Age		
– 25	–0.4246	(0.0032)
26 – 29	–0.2332	(0.0027)
30 – 34	–0.0394	(0.0026)
35 – 44	0.1728	(0.0023)
45 – 59	0.2960	(0.0024)
60 –	0.2283	(0.0060)

Cost of living area		
3	–0.0391	(0.0016)
4	–0.0183	(0.0021)
5	0.0574	(0.0020)

Education		
Degree in engineering	0.1949	(0.C
Certificate in engineering I	–0.1505	(0.C
Certificate in engineering II	–0.2327	(0.C
Degree in science	0.0828	(0.C
Other degrees	0.1056	(0.C

Industry		
Mining	0.0364	(0.0065)
Metal and engineering industry		
Iron and steel works, metal plants	0.0239	(0.0045)
Manufacture of hardware	0.0270	(0.0068)
Engineering works	–0.0339	(0.0032)
Repair works	–0.0370	(0.0171)
Shipyards	–0.0813	(0.0061)
Manufacture of electrical equipment	–0.0455	(0.0037)
Other metal industry	–0.0007	(0.0107)
Quarrying: stone, clay and glass products	0.0362	(0.0066)
Wood industry	0.0177	(0.0100)
Manufacture of pulp, paper and paper products	0.0646	(0.0059)
Printing and allied industries	–0.0473	(0.0239)
Food manufacturing industries	–0.0160	(0.0087)
Beverage and tobacco industries	–0.0095	(0.0194)
Textile industry	0.0264	(0.0100)
Leather, furs and rubber industries	–0.0093	(0.0107)
Chemical industry	0.0211	(0.0049)
Building and construction	0.0271	(0.0038)

	Sum of squares	d.f.
Total	4228.9225	34,539
Explained	2798.7530	28
Residual	1430.1695	34,511

TABLE IIIb: P64, number of observations

Total number of observations: 50,121

Age	
– 25	161
26 – 29	656
30 – 34	1,142
35 – 44	3,316
45 – 59	2,727
60 –	308

Education	
Degree in engineering	1,183
Certificate in engineering I	2,785
Certificate in engineering II	4,342

Job

Family	Level 2	3	4	5	6	7
Supervision of production	414	1,075	1,626	1,279		
Work supervision, general			260	586		
Work supervision, building and construction work				179	460	73
Technical instruction and technical service		70	281	479	86	
Other technical work	58	280	620	484		

Interaction

Age	Education: Degree in engineering	Certificate in engineering I	Certificate in engineering II
– 29	104	310	403
30 – 44	679	1,509	2,270
45 –	400	966	1,669

Age	Level 2	3	4	5	6	7
– 29		7	92	400	270	48
30 – 44	172	684	1,601	1,781	204	16
45 –	300	734	1,094	826	72	9

Age	Family: Supervision of production	Work supervision, general	Work supervision, building and construction work	Technical instruction and technical service	Other technical work
– 29	274	73	275	141	54
30 – 44	2,455	479	269	535	720
45 –	1,665	294	168	240	668

Education	Level 2	3	4	5	6	7
Degree in engineering	291	424	339	125	4	
Certificate in engineering I	116	510	1,066	983	92	18
Certificate in engineering II	65	491	1,382	1,899	450	55

Education	Family: Supervision of production	Work supervision, general	Work supervision, building and construction work	Technical instruction and technical service	Other technical work
Degree in engineering	943	10	5	55	170
Certificate in engineering I	1,710	111	105	318	541
Certificate in engineering II	1,741	725	602	543	731

TABLE IIIb:P64,1

Intercept 8.0742 (0.0017)

R = 0.8666

Age

– 25	–0.2994	(0.0130)
26 – 29	–0.1876	(0.0065)
30 – 34	–0.0880	(0.0044)
35 – 44	0.0217	(0.0021)
45 – 59	0.0690	(0.0026)
60 –	0.0379	(0.0087)

Education

Degree in engineering	0.1473	(0.0048)
Certificate in engineering I	–0.0122	(0.0025)
Certificate in engineering II	–0.0323	(0.0018)

Job

Family	Level					
	2	3	4	5	6	
Supervision of production	0.4458 (0.0085)	0.2482 (0.0047)	0.0544 (0.0036)	–0.1050 (0.0043)		
Work supervision, general			0.0209 (0.0098)	–0.1314 (0.0065)		
Work supervision, building and construction work				–0.2463 (0.0160)	–0.2783 (0.0086)	– (
Technical instruction and technical service		0.2429 (0.0193)	–0.0205 (0.0093)	–0.2157 (0.0071)	–0.3525 (0.0186)	
Other technical work	0.4088 (0.0223)	0.1982 (0.0095)	–0.0069 (0.0061)	–0.1779 (0.0073)		

	Sum of squares	d.f.
Total	778.9944	8,309
Explained	585.0480	61
Residual	193.9464	8,248

Interaction

Age	Education		
	Degree in engineering	Certificate in engineering I	Certificate in engineering I
– 29	0.0056 (0.0183)	0.0070 (0.0081)	–0.0068 (0.0071)
30 – 44	0.0096 (0.0043)	–0.0029 (0.0023)	–0.0010 (0.0018)
45 –	–0.0177 (0.0075)	0.0022 (0.0036)	0.0030 (0.0024)

Age	Level					
	2	3	4	5	6	
– 29		–0.1531 (0.0607)	–0.0575 (0.0185)	–0.0129 (0.0085)	0.0335 (0.0131)	
30 – 44	–0.0612 (0.0103)	–0.0136 (0.0042)	0.0012 (0.0022)	0.0135 (0.0023)	–0.0232 (0.0147)	–
45 –	0.0351 (0.0059)	0.0141 (0.0039)	0.0031 (0.0031)	–0.0229 (0.0043)	–0.0599 (0.0247)	–

Age	Family				
	Supervision of production	Work supervision, general	Work supervision, building and construction work	Technical instruction and technical service	Other technical
– 29	–0.0244 (0.0124)	0.0564 (0.0198)	0.0042 (0.0163)	0.0131 (0.0123)	–0.0080 (0.0227)
30 – 44	–0.0007 (0.0017)	–0.0040 (0.0047)	0.0085 (0.0128)	0.0055 (0.0042)	–0.0024 (0.0038)
45 –	0.0050 (0.0024)	–0.0074 (0.0074)	–0.0205 (0.0173)	–0.0200 (0.0086)	0.0032 (0.0042)

TABLE IIIc: P64,1

Intercept 7.8784 (0.0009)

R = 0.9171

Age

– 25	–0.2152	(0.0039)
26 – 29	–0.1185	(0.0024)
30 – 34	–0.0332	(0.0020)
35 – 44	0.0574	(0.0014)
45 – 59	0.0980	(0.0021)
60 –	0.0754	(0.0061)

Education

Degree in engineering	0.1433	(0.0028)
Certificate in engineering I	–0.0188	(0.0014)
Certificate in engineering II	–0.0384	(0.0012)
Degree in science	0.0455	(0.0080)
Other degrees	0.0705	(0.0146)

Job

Family	Level 2	3	4	5	6	7
Mathematical work	0.5233 (0.0314)	0.3750 (0.0118)	0.1348 (0.0070)	–0.0679 (0.0069)	–0.1986 (0.0115)	
Laboratory work	0.5714 (0.0117)	0.3612 (0.0063)	0.1345 (0.0035)	–0.0880 (0.0027)	–0.2140 (0.0054)	–0.2347 (0.0122
Mechanical and electrical design engineering	0.5345 (0.0111)	0.3059 (0.0049)	0.0924 (0.0027)	–0.0783 (0.0020)	–0.2219 (0.0031)	–0.2681 (0.0072
Productivity engineering	0.4903 (0.0218)	0.2581 (0.0095)	0.0793 (0.0057)	–0.0940 (0.0044)	–0.2224 (0.0082)	

	Sum of squares	d.f.
Total	1463.0585	16,266
Explained	1230.5305	80
Residual	232.5280	16,186

Interaction

Age	Education: Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in science	Other degrees
– 29	0.0302 (0.0068)	–0.0026 (0.0023)	–0.0085 (0.0027)	0.0602 (0.0216)	0.0273 (0.0555)
30 – 44	–0.0018 (0.0022)	0.0022 (0.0014)	–0.0003 (0.0010)	–0.0143 (0.0064)	0.0134 (0.0118)
45 –	–0.0228 (0.0056)	–0.0012 (0.0029)	0.0079 (0.0021)	0.0010 (0.0139)	–0.0346 (0.0235)

Age	Level 2	3	4	5	6	7
– 29	–0.3914 (0.0714)	–0.1227 (0.0288)	–0.0968 (0.0082)	–0.0081 (0.0020)	0.0248 (0.0020)	0.0312 (0.0038
30 – 44	–0.0489 (0.0082)	–0.0128 (0.0026)	0.0051 (0.0013)	0.0075 (0.0010)	–0.0195 (0.0030)	–0.0362 (0.0118
45 –	0.0428 (0.0062)	0.0205 (0.0036)	0.0134 (0.0023)	–0.0152 (0.0030)	–0.0781 (0.0078)	–0.2181 (0.0217

Education	Level					
	2	3	4	5	6	7
Degree in engineering	−0.0101 (0.0056)	−0.0084 (0.0054)	0.0027 (0.0068)	0.0378 (0.0144)	0.2153 (0.1722)	
Certificate in engineering I	0.0415 (0.0124)	0.0038 (0.0058)	−0.0012 (0.0033)	−0.0034 (0.0034)	−0.0203 (0.0228)	−0.015 (0.038
Certificate in engineering II	−0.0288 (0.0179)	0.0033 (0.0055)	0.0003 (0.0026)	−0.0007 (0.0019)	0.0022 (0.0048)	0.004 (0.012

Education	Family				
	Supervision of production	Work supervision, general	Work supervision, building and construction work	Technical instruction and technical service	Other technical work
Degree in engineering	0.0025 (0.0024)	0.0201 (0.0493)	−0.0599 (0.1536)	0.0263 (0.0210)	−0.0219 (0.0115)
Certificate in engineering I	−0.0027 (0.0021)	−0.0313 (0.0135)	0.0163 (0.0213)	0.0097 (0.0068)	0.0061 (0.0048)
Certificate in engineering II	0.0013 (0.0021)	0.0045 (0.0022)	−0.0023 (0.0039)	−0.0084 (0.0041)	0.0006 (0.0038)

TABLE III c: P64, number of observations

Total number of observations: 16,267

Age	
–25	1184
26–29	2629
30–34	3156
35–44	5699
45–59	3197
60–	402

Education	
Degree in engineering	2667
Certificate in engineering I	5608
Certificate in engineering II	7655
Degree in science	267
Other degrees	70

Job

Family	Level					
	2	3	4	5	6	7
Mathematical work	16	140	402	353	159	
Laboratory work	141	686	1430	2085	629	107
Mechanical and electrical design engineering	142	783	2162	3251	1778	315
Productivity engineering	40	190	466	750	242	

Interaction

Age	Education				
	Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in science	Other degrees
–29	503	1720	1548	37	5
30–44	1627	2730	4293	161	44
45–	537	1158	1814	69	21

Age	Level					
	2	3	4	5	6	7
–29	3	18	331	1590	1566	305
30–44	144	1030	2792	3807	994	88
45–	192	751	1337	1042	248	29

Age	Family			
	Mathematical work	Laboratory work	Mechanical and electrical design engineering	Productivity engineering
–29	276	1374	1789	334
30–44	629	2820	4409	997
45–	165	884	2233	317

Education	Level					
	2	3	4	5	6	7
Degree in engineering	230	844	1125	445	23	
Certificate in engineering I	47	436	1472	2436	1082	135
Certificate in engineering II	33	377	1721	3536	1701	287
Degree in science	22	115	112	17	1	
Other degrees	7	27	29	5	1	

Education	Family			
	Mathematical work	Laboratory work	Mechanical and electrical design engineering	Productivity engineering
Degree in engineering	448	1391	738	90
Certificate in engineering I	319	1637	2975	677
Certificate in engineering II	241	1778	4715	921
Degree in science	55	209	3	
Other degrees	7	63		

TABLE IIIc:P64,1

Intercept 7.8784 (0.0009)

R = 0.9171

Age

– 25	−0.2152	(0.0039)
26 – 29	−0.1185	(0.0024)
30 – 34	−0.0332	(0.0020)
35 – 44	0.0574	(0.0014)
45 – 59	0.0980	(0.0021)
60 –	0.0754	(0.0061)

Education

Degree in engineering	0.1433	(0.0028)
Certificate in engineering I	−0.0188	(0.0014)
Certificate in engineering II	−0.0384	(0.0012)
Degree in science	0.0455	(0.0080)
Other degrees	0.0705	(0.0146)

Job

Family	Level 2	3	4	5	6	7
Mathematical work	0.5233 (0.0314)	0.3750 (0.0118)	0.1348 (0.0070)	−0.0679 (0.0069)	−0.1986 (0.0115)	
Laboratory work	0.5714 (0.0117)	0.3612 (0.0063)	0.1345 (0.0035)	−0.0880 (0.0027)	−0.2140 (0.0054)	−0.2347 (0.0122)
Mechanical and electrical design engineering	0.5345 (0.0111)	0.3059 (0.0049)	0.0924 (0.0027)	−0.0783 (0.0020)	−0.2219 (0.0031)	−0.2681 (0.0072)
Productivity engineering	0.4903 (0.0218)	0.2581 (0.0095)	0.0793 (0.0057)	−0.0940 (0.0044)	−0.2224 (0.0082)	

	Sum of squares	d.f.
Total	1463.0585	16,266
Explained	1230.5305	80
Residual	232.5280	16,186

Interaction

Age	Education: Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in science	Other degrees
– 29	0.0302 (0.0068)	−0.0026 (0.0023)	−0.0085 (0.0027)	0.0602 (0.0216)	0.0273 (0.0555)
30 – 44	−0.0018 (0.0022)	0.0022 (0.0014)	−0.0003 (0.0010)	−0.0143 (0.0064)	0.0134 (0.0118)
45 –	−0.0228 (0.0056)	−0.0012 (0.0029)	0.0079 (0.0021)	0.0010 (0.0139)	−0.0346 (0.0235)

Age	Level 2	3	4	5	6	7
– 29	−0.3914 (0.0714)	−0.1227 (0.0288)	−0.0968 (0.0082)	−0.0081 (0.0020)	0.0248 (0.0020)	0.0312 (0.0038
30 – 44	−0.0489 (0.0082)	−0.0128 (0.0026)	0.0051 (0.0013)	0.0075 (0.0010)	−0.0195 (0.0030)	−0.0362 (0.0118
45 –	0.0428 (0.0062)	0.0205 (0.0036)	0.0134 (0.0023)	−0.0152 (0.0030)	−0.0781 (0.0078)	−0.2181 (0.0217

Age	Family			
	Mathematical work	Laboratory work	Mechanical and electrical design engineering	Productivity engineering
– 29	−0.0114 (0.0071)	0.0052 (0.0027)	−0.0048 (0.0022)	0.0134 (0.0061)
30 – 44	0.0048 (0.0032)	−0.0040 (0.0014)	0.0017 (0.0009)	0.0009 (0.0023)
45 –	0.0008 (0.0091)	0.0049 (0.0036)	0.0005 (0.0016)	−0.0170 (0.0062)

Education	Level					
	2	3	4	5	6	7
Degree in engineering	−0.0109 (0.0049)	−0.0083 (0.0033)	−0.0016 (0.0027)	0.0234 (0.0062)	0.0377 (0.0251)	
Certificate in engineering I	0.0149 (0.0180)	0.0140 (0.0054)	−0.0020 (0.0024)	−0.0066 (0.0015)	0.0021 (0.0030)	0.0306 (0.0089)
Certificate in engineering II	0.0114 (0.0212)	−0.0035 (0.0059)	−0.0001 (0.0022)	0.0015 (0.0011)	−0.0019 (0.0019)	−0.0144 (0.0042)
Degree in science	0.0533 (0.0260)	−0.0005 (0.0095)	−0.0144 (0.0096)	0.0224 (0.0297)	0.1108 (0.1203)	
Other degrees	−0.0367 (0.0439)	0.0152 (0.0185)	0.0088 (0.0177)	−0.0145 (0.0532)	−0.0294 (0.1219)	

Education	Family			
	Mathematical work	Laboratory work	Mechanical and electrical design engineering	Productivity engineering
Degree in engineering	−0.0143 (0.0051)	−0.0139 (0.0024)	0.0319 (0.0042)	0.0240 (0.0129)
Certificate in engineering I	0.0103 (0.0059)	0.0158 (0.0021)	−0.0078 (0.0012)	−0.0085 (0.0034)
Certificate in engineering II	0.0049 (0.0076)	−0.0026 (0.0021)	−0.0000 (0.0008)	0.0039 (0.0026)
Degree in science	0.0377 (0.0155)	−0.0098 (0.0041)	−0.0057 (0.0696)	
Other degrees	−0.0178 (0.0456)	0.0020 (0.0051)		

Table IV:P64, number of observations

Total number of observations: 3,160

Age	
−25	65
26–29	282
30–34	565
35–44	1369
45–59	758
60–	121

Cost of living area	
3	1644
4	548
5	968

Education	
Degree in engineering	480
Certificate in engineering I	958
Certificate in engineering II	1134
Degree in business & commerce	185
Certificate in commerce	403

Job	
Chief marketing manager, chief market development manager, chief sales manager (8102)	180
Marketing manager, market development manager, sales manager (8103)	526
First market development planner, first salesman, first sales engineer (8104)	1112
Market planner, salesman, sales engineer (8105)	1295
First administrative rationalization officer (9454)	47

Industry	
Metal and engineering industry	2550
Quarrying; stone, clay and glass products	119
Manufacture of pulp, paper and paper products	117
Textile industry	51
Chemical industry	171
Building and construction	152

Table IV:P64,1

Intercept 8.0691 (0.0027) R = 0.8815

Age		Cost of living area		Education	
–25	–0.3287	3	–0.0089	Degree in engineering	0.1100
26–29	–0.1920 (0.0090)	4	–0.0134 (0.0060)	Certificate in engineering I	–0.0095 (0.0041)
30–34	–0.0721 (0.0060)	5	0.0226 (0.0042)	Certificate in engineering II	–0.0316 (0.0039)
35–44	0.0247 (0.0031)			Degree in business & economics	0.0507 (0.0114)
45–59	0.1004 (0.0051)			Certificate in commerce	-0.0430 (0.0077)
69–	0.0525 (0.0137)				

Job		Industry	
Chief marketing manager, chief market development manager, chief sales manager (8102)	0.4964	Metal and engineering industry	–0.0047
Marketing manager, market development manager, sales manager (8103)	0.2771 (0.0065)	Quarrying: stone, clay and glass products	0.0767 (0.0137)
First market development planner, first salesman, first sales engineer (8104)	0.0196 (0.0037)	Manufacture of pulp, paper and paper products	0.0160 (0.0147)
Market planner, salesman, sales engineer (8105)	–0.1965 (0.0038)	Textile industry	0.0560 (0.0212)
First administrative rationalization officer (9454)	–0.0518 (0.0221)	Chemical industry	0.0196 (0.0114)
		Building and construction	–0.0344 (0.0122)

	Sum of squares	d.f.
Total	320.298878	3159
Explained	248.878386	20
Residual	71.420492	

Table IV:P64,3

Intercept 8.1459 (0.0083) R = 0.8815

Age		Cost of living area		Education	
–25	–0.2595 (0.0162)	3	–0.0090 (0.0038)	Degree in engineering	0.0947 (0.0066)
26–29	–0.1228 (0.0089)	4	–0.0135 (0.0048)	Certificate in engineering I	–0.0248 (0.0053)
30–34	–0.0029 (0.0069)	5	0.0225 (0.0042)	Certificate in engineering II	–0.0469 (0.0054)
35–44	0.0939 (0.0057)			Degree in business & economics	0.0354 (0.0097)
45–59	0.1696 (0.0067)			Certificate in commerce	–0.0584 (0.0072)
60–	0.1217 (0.0124)				

Job	
Chief marketing manager, chief market development manager, chief sales manager (8102)	0.3874 (0.0106)
Marketing manager, market development manager, sales manager (8103)	0.1682 (0.0075)
First market development planner, first salesman, first sales engineer (8104)	–0.0894 (0.0064)
Market planner, salesman, sales engineer (8105)	–0.3055 (0.0069)
First administrative rationalization officer (9454)	–0.1608 (0.0181)

Industry	
Metal and engineering industry	–0.0262 (0.0064)
Quarrying: stone, clay and glass products	0.0551 (0.0127)
Manufacture of pulp, paper and paper products	–0.0055 (0.0134)
Textile industry	0.0345 (0.0183)
Chemical industry	–0.0019 (0.0110)
Building and construction	–0.0559 (0.0119)

	Sum of squares	d.f.
Total	320.2989	3159
Explained	248.8801	20
Residual	71.4188	3139

TABLE VII:P64,1

Intercept 8.0691 (0.0027)

R = 0.8959

Age		
– 25	–0.3346	(0.0195)
26 – 29	–0.1931	(0.0092)
30 – 34	–0.0739	(0.0061)
35 – 44	0.0261	(0.0032)
45 – 59	0.1005	(0.0051)
60 –	0.0503	(0.0139)

Cost of living area		
3	–0.0096	(0.0027)
4	–0.0120	(0.0060)
5	0.0230	(0.0042)

Education		
Degree in engineering	0.1103	(0.0068)
Certificate in engineering I	–0.0115	(0.0042)
Certificate in engineering II	–0.0321	(0.0039)
Degree in business & economics	0.0625	(0.0117)
Certificate in commerce	–0.0423	(0.0077)

Job level		
8102	0.4920	(0.0120)
8103	0.2765	(0.0065)
8104	0.0173	(0.0038)
8105	–0.1934	(0.0039)
9454	–0.0584	(0.0222)

Industry		
Metal and engineering industry	–0.0054	(0.0014)
Quarrying stone, clay and glass products	0.0832	(0.0138)
Manufacture of pulp, paper and paper products	0.0167	(0.0149)
Textile industry	0.0592	(0.0217)
Chemical industry	0.0223	(0.0116)
Building and construction	–0.0321	(0.0123)

	Sum of squares	d.f.
Total	320.2989	3,159
Explained	251.3594	70
Residual	68.9395	3,089

Interaction

Age	Cost of living area		
	3	4	5
Number of observations			
– 29	177	61	109
30 – 44	990	362	582
45 –	477	125	277
Estimates			
– 29	–0.0084	–0.0112	0.0199
	(0.0084)	(0.0183)	(0.0127)
30 – 44	–0.0031	0.0004	0.0050
	(0.0022)	(0.0043)	(0.0035)
45 –	0.0095	0.0044	–0.0183
	(0.0045)	(0.0116)	(0.0070)

Age	Education: Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce
Number of observations					
–29	27	136	106	32	46
30–44	279	567	720	124	244
45–	174	255	308	29	113
Estimates					
–29	0.0322	-0.0097	0.0031	0.0150	-0.0078
	(0.0318)	(0.0105)	(0.0130)	(0.0332)	(0.0214)
30–44	0.0068	-0.0027	-0.0016	0.0004	0.0030
	(0.0059)	(0.0035)	(0.0030)	(0.0089)	(0.0059)
45–	-0.0159	0.0112	0.0027	-0.0182	-0.0034
	(0.0093)	(0.0075)	(0.0068)	(0.0294)	(0.0122)

Age	Job: 8102	8103	8104	8105	9454
Number of observations					
–29	0	2	43	296	6
30–44	83	268	734	810	39
45–	97	256	335	189	2
Estimates					
–29	–	-0.1839	-0.0761	0.0131	-0.0380
		(0.1159)	(0.0241)	(0.0039)	(0.0657)
30–44	-0.0290	-0.0255	0.0104	0.0012	0.0175
	(0.0132)	(0.0059)	(0.0027)	(0.0024)	(0.0110)
45–	0.0248	0.0281	-0.0129	-0.0255	-0.2265
	(0.0113)	(0.0062)	(0.0055)	(0.0092)	(0.1089)

Cost of living area	Education: Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce
Number of observations					
3	284	530	439	87	304
4	45	152	242	57	52
5	151	276	453	41	47
Estimates					
3	0.0102	0.0010	0.0003	-0.0007	-0.0115
	(0.0057)	(0.0037)	(0.0047)	(0.0126)	(0.0042)
4	-0.0126	0.0094	-0.0200	0.0306	0.0430
	(0.0217)	(0.0096)	(0.0070)	(0.0172)	(0.0189)
5	-0.0154	-0.0070	0.0104	-0.0410	0.0265
	(0.0102)	(0.0065)	(0.0044)	(0.0216)	(0.0203)

Cost of living area	Job: 8102	8103	8104	8105	9454
Number of observations					
3	106	261	596	658	23
4	32	80	199	223	14
5	42	185	317	414	10
Estimates					
3	-0.0126	0.0061	0.0029	-0.0035	0.0126
	(0.0101)	(0.0066)	(0.0036)	(0.0038)	(0.0233)
4	0.0101	-0.0091	-0.0098	0.0130	-0.0391
	(0.0261)	(0.0154)	(0.0081)	(0.0083)	(0.0347)
5	0.0241	-0.0047	0.0007	-0.0015	0.0257
	(0.0219)	(0.0089)	(0.0061)	(0.0055)	(0.0457)

Education	Job				
	8102	8103	8104	8105	9454
Number of observations					
Degree in engineering	67	153	211	44	5
Certificate in engineering I	27	131	327	463	10
Certificate in engineering II	18	130	382	594	10
Degree in business & economics	39	47	61	25	13
Certificate in commerce	29	65	131	169	9
Estimates					
Degree in engineering	-0.0200	-0.0079	0.0035	0.0380	0.0277
	(0.0142)	(0.0091)	(0.0072)	(0.0231)	(0.0650)
Certificate in engineering I	-0.0117	0.0114	-0.0003	-0.0024	0.0051
	(0.0274)	(0.0110)	(0.0059)	(0.0046)	(0.0432)
Certificate in engineering II	-0.0099	-0.0002	0.0041	-0.0032	0.0515
	(0.0344)	(0.0111)	(0.0053)	(0.0037)	(0.0450)
Degree in business & economics	0.0198	-0.0137	-0.0148	0.0160	0.0288
	(0.0222)	(0.0198)	(0.0165)	(0.0364)	(0.0380)
Certificate in commerce	0.0367	0.0060	-0.0101	0.0056	-0.1198
	(0.0258)	(0.0168)	(0.0105)	(0.0090)	(0.0455)

TABLE VIIIa: P60, number of observations

Total number of observations: 5,850

Age		Cost of living area		Education	
– 25	374	3	3,112	Degree in engineering	512
26 – 29	715	4	966	Certificate in engineering I	1,275
30 – 34	1,150	5	1,772	Certificate in engineering II	1,751
35 – 44	2,248			Degree in business & economics	625
45 – 59	1,200			Certificate in commerce	1,687
60 –	163				

Job

Family	Level					
	2	3	4	5	6	7
Marketing	169	455	922	967		
Sales Activity through calls on customers		49	122	240		
Tender calculations		24	122	297		
Handling of orders			56	150		
Advertising		26	45	28		
Purchasing	23	85	151	166		
Financial administration	110	163	206	150	442	210
Accounting, budget and cashier work	15	80	128	132		
Internal auditing		16	20	21		
Office rationalization		25	41	29		
Administrative rationalization		0	5	6		

Industry

Mining	66
Metal and engineering industry	
Iron and steel works, metal plants	347
Manufacture of hardware	184
Engineering works	1,734
Repair works	109
Shipyards	106
Manufacture of electrical equipment	1,184
Other metal industry	85
Quarrying: stone, clay and glass products	208
Wood industry	116
Manufacture of pulp, paper and paper products	254
Printing and allied industries	35
Food manufacturing industries	122
Beverage and tobacco industries	31
Textile industry	177
Leather, furs and rubber industries	134
Chemical industry	306
Building and construction	652

TABLE VIIIa:P60,1

Intercept 7.6553 (0.0023) R = 0.8796

Age		
− 25	−0.2837	(0.0100)
26 − 29	−0.1602	(0.0066)
30 − 34	−0.0431	(0.0048)
35 − 44	0.0490	(0.0031)
45 − 59	0.1202	(0.0048)
60 −	0.9075	(0.0139)

Cost of living area		
3	−0.0261	(0.0024)
4	0.0035	(0.0054)
5	0.0440	(0.0038)

Education		
Degree in engineering	0.1388	(0.0085)
Certificate in engineering I	0.0054	(0.0048)
Certificate in engineering II	−0.0325	(0.0042)
Degree in business & economics	0.0519	(0.0079)
Certificate in commerce	−0.0317	(0.0048)

Job

Family	Level 2	3	4	5	6	
Marketing	0.5548	0.3614	0.0806	−0.01178		
	(0.0140)	(0.0085)	(0.0057)	(0.0056)		
Sales Activity through calls on customers		0.3026	0.1253	−0.0200		
		(0.0254)	(0.0161)	(0.0155)		
Tender calculations		0.2889	0.0342	−0.1706		
		(0.0363)	(0.0164)	(0.0107)		
Handling of orders			−0.0349	−0.2029		
			(0.0237)	(0.0147)		
Advertising		0.3216	0.0797	−0.2220		
		(0.0348)	(0.0264)	(0.0335)		
Purchasing	0.5373	0.2443	−0.0430	−0.2074		
	(0.0373)	(0.0192)	(0.0143)	(0.0164)		
Financial administration	0.4112	0.2396	0.0074	−0.1776	−0.3246	−0
	(0.0177)	(0.0143)	(0.0128)	(0.0149)	(0.0088)	(0.
Accounting, budget and cashier work	0.5527	0.2322	0.0091	−0.2269		
	(0.0463)	(0.0203)	(0.0158)	(0.0156)		
Internal auditing		0.3840	0.0714	−0.1735		
		(0.0449)	(0.0399)	(0.0389)		
Office rationalization		0.2838	0.0178	−0.2025		
		(0.0358)	(0.0278)	(0.0332)		
Administrative rationalization		−	0.2399	−0.0550		
		−	(0.0795)	(0.0726)		

Industry		
Mining	0.0886	(0.0221)
Metal and engineering industry		
Iron and steel works, metal plants	0.0397	(0.0097)
Manufacture of hardware	−0.0065	(0.0130)
Engineering works	0.0041	(0.0037)
Repair works	0.0165	(0.0170)
Shipyards	0.0617	(0.0175)
Manufacture of electrical equipment	−0.0352	(0.0051)
Other metal industry	−0.0100	(0.0192)
Quarrying: stone, clay and glass products	0.0224	(0.0122)
Wood industry	−0.0343	(0.0166)
Manufacture of pulp, paper and paper products	0.0359	(0.0114)
Printing and allied industries	−0.0384	(0.0302)
Food manufacturing industries	−0.0290	(0.0163)
Beverage and tobacco industries	0.0114	(0.0322)
Textile industry	0.0304	(0.0134)
Leather, furs and rubber industries	0.0003	(0.0155)
Chemical industry	0.0279	(0.0101)
Building and construction	−0.0162	(0.0071)

	Sum of squares	d.f.
Total	801.7369	5,849
Explained	620.3099	64
Residual	181.4271	5,785

TABLE VIIIa:P60,3

Intercept 7.7099 (0.0080)

R = 0.8803

Age			Cost of living area			Education		
– 25	−0.2454	(0.0090)	3	−0.0334	(0.0034)	Degree in engineering	0.1121	(0.0074)
26 – 29	−0.1237	(0.0066)	4	−0.0032	(0.0044)	Certificate in engineering I	−0.0214	(0.0051)
30 – 34	−0.0066	(0.0055)	5	0.0366	(0.0038)	Certificate in engineering II	−0.0592	(0.0050)
35 – 44	0.0855	(0.0048)				Degree in business & economics	0.0256	(0.0072)
45 – 59	0.1564	(0.0056)				Certificate in commerce	−0.0570	(0.0054)
60 –	0.1338	(0.0120)						

Job

Family	Level 2	3	4	5	6	7
Marketing	0.5068 (0.0151)	0.3136 (0.0108)	0.0328 (0.0090)	−0.1657 (0.0090)		
Sales Activity through calls on customers		0.2545 (0.0256)	0.0775 (0.0172)	−0.0677 (0.0133)		
Tender calculations		0.2414 (0.0360)	−0.0133 (0.0176)	−0.2184 (0.0128)		
Handling of orders			−0.0831 (0.0241)	−0.2507 (0.0160)		
Advertising		0.2736 (0.0345)	0.0315 (0.0266)	−0.2703 (0.0332)		
Purchasing	0.4892 (0.0368)	0.1961 (0.0199)	−0.0910 (0.0156)	−0.2557 (0.0175)		
Financial administration	0.3631 (0.0182)	0.1912 (0.0154)	−0.0412 (0.0141)	−0.2263 (0.0160)	−0.3736 (0.0111)	−0.5297 (0.0145)
Accounting, budget and cashier work	0.5049 (0.0454)	0.1840 (0.0207)	−0.0394 (0.0167)	−0.2758 (0.0166)		
Internal auditing		0.3359 (0.0440)	0.0230 (0.0392)	−0.2222 (0.0383)		
Office rationalization		0.2354 (0.0353)	−0.0303 (0.0278)	−0.2515 (0.0329)		
Administrative rationalization		–	0.1922 (0.0776)	−0.1029 (0.0709)	−0.2415 (0.1223)	−0.6962 (0.1222)

Industry

Mining	0.0790	(0.0212)
Metal and engineering industry		
Iron and steel works, metal plants	0.0318	(0.0102)
Manufacture of hardware	−0.0160	(0.0130)
Engineering works	−0.0049	(0.0060)
Repair works	0.0070	(0.0167)
Shipyards	0.0523	(0.0171)
Manufacture of electrical equipment	−0.0444	(0.0070)
Other metal industry	−0.0195	(0.0186)
Quarrying: stone, clay and glass products	0.0128	(0.0123)
Wood industry	−0.0443	(0.0161)
Manufacture of pulp, paper and paper products	0.0258	(0.0115)
Printing and allied industries	−0.0480	(0.0288)
Food manufacturing industries	−0.0370	(0.0159)
Beverage and tobacco industries	0.0018	(0.0306)
Textile industry	0.0207	(0.0133)
Leather, furs and rubber industries	−0.0096	(0.0152)
Chemical industry	0.0182	(0.0104)
Building and construction	−0.0256	(0.0083)

	Sum of squares	d.f.
Total	801.7369	5,849
Explained	621.2379	66
Residual	180.4990	5,783

TABLE VIIIa: P64, number of observations

Total number of observations: 7,381

Age

– 25	328
26 – 29	807
30 – 34	1,307
35 – 44	2,876
45 – 59	1,799
60 –	264

Cost of living area

3	4,022
4	1,332
5	2,027

Education

Degree in engineering	578
Certificate in engineering I	1,690
Certificate in engineering II	2,240
Degree in business & economics	708
Certificate in commerce	2,165

Job

Family	Level					
	2	3	4	5	6	7
Marketing	197	561	1,163	1,323		
Sales Activity through calls on customers		57	141	255		
Tender calculations		52	160	329		
Handling of orders			66	149		
Advertising		29	59	34		
Purchasing	31	99	217	162		
Financial administration	124	198	275	162	506	192
Accounting, budget and cashier work	33	97	187	171		
Internal auditing		18	36	26		
Office rationalization		34	54	55		
Administrative rationalization		16	36	58	18	1

Industry

Mining	63
Metal and engineering industry	
Iron and steel works, metal plants	472
Manufacture of hardware	228
Engineering works	2,272
Repair works	162
Shipyards	146
Manufacture of electrical equipment	1,442
Other metal industry	109
Quarrying: stone, clay and glass products	304
Wood industry	130
Manufacture of pulp, paper and paper products	338
Printing and allied industries	65
Food manufacturing industries	118
Beverage and tobacco industries	56
Textile industry	186
Leather, furs and rubber industries	171
Chemical industry	389
Building and construction	730

TABLE VIIIa:P64,1

Intercept 7.9829 (0.0018)

R = 0.8902

Age

– 25	–0.2785	(0.0090)
26 – 29	–0.1664	(0.0055)
30 – 34	–0.0557	(0.0040)
35 – 44	0.0376	(0.0023)
45 – 59	0.0984	(0.0034)
60 –	0.0498	(0.0095)

Cost of living area

3	–0.0169	(0.0018)
4	–0.0046	(0.0040)
5	0.0356	(0.0032)

Education

Degree in engineering	0.1354	(0.0069)
Certificate in engineering I	0.0022	(0.0036)
Certificate in engineering II	–0.0216	(0.0032)
Degree in business & economics	0.0581	(0.0066)
Certificate in commerce	–0.0345	(0.0036)

Job

Family	Level					
	2	3	4	5	6	7
Marketing	0.5537 (0.0114)	0.3481 (0.0066)	0.0904 (0.0044)	–0.1328 (0.0042)		
Sales Activity through calls on customers		0.2623 (0.0206)	0.1257 (0.0131)	–0.0188 (0.0098)		
Tender calculations		0.2506 (0.0218)	0.0317 (0.0127)	–0.1560 (0.0090)		
Handling of orders			–0.0405 (0.0192)	–0.2182 (0.0128)		
Advertising		0.3087 (0.0289)	0.0281 (0.0202)	–0.1772 (0.0266)		
Purchasing	0.5062 (0.0280)	8.2332 (0.0156)	–0.0347 (0.0105)	–0.2027 (0.0122)		
Financial administration	0.4042 (0.0146)	0.1944 (0.0114)	–0.0151 (0.0099)	–0.1843 (0.0126)	–0.3376 (0.0071)	–0.4712 (0.0115)
Accounting, budget and cashier work	0.5115 (0.0274)	0.2322 (0.0160)	–0.0079 (0.0115)	–0.2002 (0.0120)		
Internal auditing		0.3004 (0.0370)	0.0615 (0.0260)	–0.2014 (0.0306)		
Office rationalization		0.2469 (0.0268)	0.0107 (0.0211)	–0.1764 (0.0210)		
Administrative rationalization		0.3082 (0.0390)	0.1328 (0.0259)	–0.0827 (0.0204)	–0.2502 (0.0368)	–0.5485 (0.1552)

Industry

Mining	0.0595	(0.0198)
Metal and engineering industry		
Iron and steel works, metal plants	–0.0022	(0.0072)
Manufacture of hardware	0.0027	(0.0102)
Engineering works	–0.0012	(0.0028)
Repair works	0.0070	(0.0123)
Shipyards	0.0100	(0.0130)
Manufacture of electrical equipment	–0.0298	(0.0041)
Other metal industry	–0.0246	(0.0148)
Quarrying: stone, clay and glass products	0.0375	(0.0088)
Wood industry	–0.0351	(0.0136)
Manufacture of pulp, paper and paper products	0.0121	(0.0086)
Printing and allied industries	–0.0299	(0.0195)
Food manufacturing industries	–0.0243	(0.0145)
Beverage and tobacco industries	0.0439	(0.0210)
Textile industry	0.0122	(0.0114)
Leather, furs and rubber industries	–0.0030	(0.0120)
Chemical industry	0.0183	(0.0078)
Building and construction	0.0343	(0.0061)

TABLE VIIIa: P64,3

Intercept 8.0106 (0.0061) R = 0.8901

Age		
– 25	−0.2261	(0.0079)
26 – 29	−0.1138	(0.0053)
30 – 34	−0.0033	(0.0045)
35 – 44	0.0901	(0.0037)
45 – 59	0.1508	(0.0042)
60 –	0.1022	(0.0083)

Cost of living area		
3	−0.0219	(0.0027)
4	−0.0096	(0.0033)
5	0.0315	(0.0030)

Education		
Degree in engineering	0.1075	(0.0060)
Certificate in engineering I	−0.0258	(0.0040)
Certificate in engineering II	−0.0496	(0.0039)
Degree in business & economics	0.0302	(0.0059)
Certificate of commerce	−0.0624	(0.0042)

Job

Family	Level 2	3	4	5	6	7
Marketing	0.5113	0.3057	0.0480	−0.1753		
	(0.0122)	(0.0084)	(0.0070)	(0.0070)		
Sales Activity through calls on customers		0.2199	0.0833	−0.0612		
		(0.0207)	(0.0139)	(0.0111)		
Tender calculations		0.2082	−0.0107	−0.1984		
		(0.0219)	(0.0136)	(0.0104)		
Handling of orders			−0.0829	−0.2596		
			(0.0195)	(0.0137)		
Advertising		0.2663	−0.0143	−0.2196		
		(0.0286)	(0.0204)	(0.0265)		
Purchasing	0.4638	0.1908	−0.0771	−0.2452		
	(0.0277)	(0.0161)	(0.0115)	(0.0131)		
Financial administration	0.3619	0.1520	−0.0575	−0.2258	−0.3700	−0.51
	(0.0150)	(0.0121)	(0.0109)	(0.0133)	(0.0087)	(0.01
Accounting, budget and cashier work	0.4691	0.1898	−0.0504	−0.2427		
	(0.0271)	(0.0163)	(0.0123)	(0.0129)		
Internal auditing		0.2580	0.0191	−0.2439		
		(0.0364)	(0.0258)	(0.0303)		
Office rationalization		0.2045	−0.0317	−0.2189		
		(0.0266)	(0.0213)	(0.0212)		
Administrative rationalization		0.2658	0.0903	−0.1251	−0.2928	−0.59
		(0.0383)	(0.0258)	(0.0206)	(0.0363)	(0.15

Industry		
Mining	0.0547	(0.0190)
Metal and engineering industry		
Iron and steel works, metal plants	−0.0070	(0.0076)
Manufacture of hardware	−0.0022	(0.0102)
Engineering works	−0.0059	(0.0046)
Repair works	0.0020	(0.0121)
Shipyards	0.0052	(0.0128)
Manufacture of electrical equipment	−0.0346	(0.0054)
Other metal industry	−0.0294	(0.0144)
Quarrying: stone, clay and glass products	0.0326	(0.0090)
Wood industry	−0.0400	(0.0132)
Manufacture of pulp, paper and paper products	0.0072	(0.0087)
Printing and allied industries	−0.0347	(0.0187)
Food manufacturing industries	−0.0292	(0.0140)
Beverage and tobacco industries	0.0390	(0.0200)
Textile industry	0.0073	(0.0112)
Leather, furs and rubber industries	−0.0080	(0.0118)
Chemical industry	0.0135	(0.0080)
Building and construction	0.0294	(0.0069)

	Sum of squares	d.f.
Total	845.8147	7,380
Explained	670.1647	67
Residual	175.6500	7,313

TABLE VIIIa:H68, number of observations

Total number of observations: 11,820

Age

– 25	1,058
26 – 29	1,739
30 – 34	2,108
35 – 44	3,542
45 – 59	2,999
60 –	374

Cost of living area

3	6,453
4	2,088
5	3,279

Education

Degree in engineering	823
Certificate in engineering I	3,068
Certificate in engineering II	3,932
Degree in business & economics	854
Certificate in commerce	3,143

Job

Family	Level 2	3	4	5	6	7
PR work	1	11	17			
Marketing	184	758	1,526	2,035	391	
Sales Activity through calls on customers		66	214	405	70	
Tender calculations		76	240	505	351	
Handling of orders		23	100	265	161	
Advertising		26	54	86		
Purchasing	44	149	270	245	119	
Financial administration	137	197	353	232		
Accounting, budget and cashier work	36	140	315	291		
Internal auditing	11	24	28	36		
Administrative rationalization	13	42	66	131	28	
Supervision of data processing work	4	9	13	5		
Systems work, programming work		35	122	271	93	
Computer operation work		5	7	2	6	15
General office work			20	20	350	381

Industry

Mining	56
Metal and engineering industry	
Iron and steel works, metal plants	748
Manufacture of hardware	460
Engineering works	3,514
Repair works	235
Shipyards	217
Manufacture of electrical equipment	2,214
Other metal industry	204
Quarrying: stone, clay and glass products	460
Wood industry	230
Manufacture of pulp, paper and paper products	424
Printing and allied industries	100
Food manufacturing industries	257
Beverage and tobacco industries	37
Textile industry	244
Leather, furs and rubber industries	190
Chemical industry	653
Building and construction	1,577

TABLE VIIIa: H68,3

Intercept 8.2630 (.0049)

R = .9103

Age

– 25	–.2511
26 – 29	–.1216
30 – 34	–.063
35 – 44	.0934
45 – 59	.1546
60 –	.1309

Cost of living area

3	–.0325	(.0020)
4	.0000	(.0025)
5	.0325	(.0023)

Education

Degree in engineering	.1086	(.0046)
Certificate in engineering I	–.0280	(.0029)
Certificate in engineering II	–.0504	(.0029)
Degree in business & economics	.0247	(.0048)
Certificate in commerce	–.0548	(.0031)

Job

Family	Level 2	3	4	5	6	[illegible]
PR work	[.3071 (.0732)]	.2133 (.0437)	.0163 (.0548)			
Marketing	.5234 (.0115)	.3023 (.0068)	.0678 (.0056)	–.1374 (.0054)	–.3305 (.0088)	
Sales Activity through calls on customers		.2978 (.0183)	.0837 (.0108)	–.0966 (.0083)	–.2569 (.0178)	
Tender calculations		.2042 (.0174)	–.0158 (.0106)	–.2021 (.0081)	–.3711 (.0094)	
Handling of orders		.2610 (.0304)	–.0578 (.0151!	–.2496 (.0098)	–.3870 (.0122)	
Advertising		.2290 (.0286)	.0184 (.0200)	–.1627 (.0161)		
Purchasing	.4632 (.0222)	.1659 (.0125)	–.0570 (.0097)	–.2127 (.0102)	–.3674 (.0140)	
Financial administration	.4018 (.0134)	.1693 (.0113)	–.0436 (.0091)	–.2617 (.0107)		
Accounting, budget and cashier work	.3615 (.0246)	.1780 (.0130)	–.0347 (.0092)	–.2266 (.0095)		
Internal auditing	.3389 (.0044)	.1984 (.0298)	–.0431 (.0276)	–.1968 (.0245)		
Administrative rationalization	.4446 (.0403)	.2506 (.0227)	.0207 (.0182)	–.1443 (.0133)	–.3288 (.0277)	
Supervision of data processing work	[.3579 (.0723)]	.2056 (.0483)	.0635 (.0402)	–.0681 (.0647)		
Systems work, programming work		.2388 (.0248)	.0446 (.0138)	–.1505 (.0099)	–.3123 (.0158)	
Computer operation work		.2202 (.0646)	.0297 (.0547)	–.0190 (.0687)	–.3027 (.0591)	–.54 (.03
General office work			--.0065 (.0325)	–.1683 (.0325)	–.4021 (.0089)	–.51 (.00

Industry

Mining	0.0144	(0.0189)
Metal and engineering industry		
Iron and steel works metal plants	–0.0221	(0.0058)
Manufacture of hardware	–0.0050	(0.0070)
Engineering works	–0.0146	(0.0036)
Repair works	0.0231	(0.0096)
Shipyards	–0.0068	(0.0099)
Manufacture of electrical equipment	–0.0214	(0.0043)
Other metal industry	–0.0271	(0.0100)

Quarrying: stone, clay and glass products	0.0106	(0.0070)
Wood industry	−0.0021	(0.0095)
Manufacture of pulp, paper and paper products	−0.0110	(0.0073)
Printing and allied industries	0.0281	(0.0142)
Food manufacturing industries	−0.0037	(0.0091)
Beverage and tobacco industries	0.0107	(0.0230)
Textile industry	0.0089	(0.0093)
Leather furs and rubber industries	−0.0246	(0.0104)
Chemical industry	−0.0003	(0.0060)
Building and construction	0.0209	(0.0050)

	Sum of squares	d.f.
Total	1469.0762	11,819
Explained	1217.4018	87
Residual	251.6744	11,732

TABLE VIIIe: P64, number of observations

Total number of observations: 5,223

Age	
–25	119
26–29	451
30–34	900
35–44	2235
45–59	1329
60–	189

Education	
Degree in engineering	538
Certificate in engineering I	1504
Certificate in engineering II	2039
Degree in business & economics	289
Certificate in commerce	853

Job

Family	Level			
	2	3	4	5
Marketing	198	577	1185	1351
Sales activities through calls on customers		62	149	285
Tender calculations		52	162	332
Handling of orders			66	151
Advertising		33	60	34
Purchasing	31	104	225	166

Interaction

Age	Education				
	Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce
–29	29	221	191	33	96
30–44	306	875	1258	190	506
45–	203	408	590	66	251

Age	Level			
	2	3	4	5
–29	0	9	73	488
30–44	102	417	1183	1433
45–	127	402	591	398

Age	Family					
	Marketing	Sales activities through calls on customers	Tender calculations	Handling of orders	Advertising	Purching
–29	356	41	75	35	8	55
30–44	2020	333	303	113	82	284
45–	935	122	168	69	37	187

Education	Level			
	2	3	4	5
Degree in engineering	76	180	231	51
Certificate in engineering I	32	201	533	738
Certificate in engineering II	20	201	666	1152
Degree in business & economics	58	90	105	36
Certificate in commerce	43	156	312	342

Education	Family					
	Marketing	Sales activities through calls on customers	Tender calculations	Handling of orders	Advertising	Purchasing
Degree in engineering	484	8	21	5	3	17
Certificate in enginering I	974	131	180	71	32	116
Certificate in enginering II	1161	232	339	95	30	182
Degree in business & economics	210	23	1	3	26	26
Certificate in commerce	482	102	5	43	36	185

TABLE VIIIe:P64,1

Intercept 8.0380 (0.0022)

R = 0.8641

Age

– 25	−0.3380	(0.0148)
26 – 29	−0.1922	(0.0075)
30 – 34	−0.0694	(0.0050)
35 – 44	0.0263	(0.0026)
45 – 59	0.0920	(0.0039)
60 –	0.0435	(0.0115)

Education

Degree in engineering	0.1178	(0.0069)
Certificate in engineering I	−0.0053	(0.0035)
Certificate in engineering II	−0.0218	(0.0029)
Degree in business & economics	0.0501	(0.0095)
Certificate in commerce	−0.0297	(0.0051)

Job

Family	Level			
	2	3	4	5
Marketing	0.5156 (0.0119)	0.3090 (0.0067)	0.0456 (0.0042)	−0.1727 (0.0041)
Sales activities through calls on customers		0.2239 (0.0021)	0.0802 (0.0129)	−0.0555 (0.0094)
Tender calculations		0.2327 (0.0234)	0.0211 (0.0126)	−0.1821 (0.0089)
Handling of orders			−0.0913 (0.0202)	−0.2587 (0.0129)
Advertising		0.2662 (0.0294)	−0.0060 (0.0204)	−0.2229 (0.0282)
Purchasing	0.4385 (0.0314)	0.1859 (0.0162)	−0.0716 (0.0106)	−0.2335 (0.0133)

Interaction

Age	Education				
	Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce
– 29	0.0238 (0.0312)	−0.0053 (0.0084)	0.0059 (0.0096)	0.0219 (0.0305)	−0.0143 (0.0151)
30 – 44	0.0047 (0.0060)	−0.0002 (0.0030)	0.0025 (0.0023)	0.0024 (0.0070)	−0.0096 (0.0043)
45 –	−0.0104 (0.0091)	0.0032 (0.0061)	−0.0072 (0.0048)	−0.0177 (0.0018)	0.0248 (0.0083)

Age	Level			
	2	3	4	5
– 29		−0.1201 (0.0539)	−0.0525 (0.0186)	0.0101 (0.0030)
30 – 44	−0.0360 (0.0124)	−0.0222 (0.0051)	0.0062 (0.0022)	0.0039 (0.0019)
45 –	0.0289 (0.0099)	0.0257 (0.0053)	−0.0060 (0.0042)	−0.0263 (0.0064)

Age	Family Marketing	Sales activities through calls on customers	Tender calculations	Handling of orders	Advertising	Purchasing
– 29	−0.0001 (0.0052)	−0.0399 (0.0236)	0.0136 (0.0170)	0.0391 (0.0246)	−0.0404 (0.0552)	−0.0074 (0.0213)
30 – 44	−0.0023 (0.0014)	−0.0075 (0.0048)	0.0175 (0.0059)	−0.0039 (0.0101)	0.0007 (0.0111)	0.0080 (0.0062)
45 –	0.0050 (0.0029)	0.0338 (0.0122)	−0.0376 (0.0102)	−0.0134 (0.0159)	0.0071 (0.0238)	−0.0099 (0.0096)

Education	Level 2	3	4	5
Degree in engineering	−0.0171 (0.0142)	−0.0113 (0.0090)	0.0047 (0.0074)	0.0440 (0.0220)
Certificate in engineering I	−0.0112 (0.0260)	0.0016 (0.0093)	0.0028 (0.0048)	−0.0020 (0.0038)
Certificate in engineering II	0.0145 (0.0333)	−0.0046 (0.0093)	−0.0026 (0.0041)	0.0021 (0.0025)
Degree in business & economics	−0.0100 (0.0176)	0.0024 (0.0137)	−0.0015 (0.0122)	0.0144 (0.0280)
Certificate in commerce	0.0453 (0.0219)	0.0156 (0.0108)	−0.0021 (0.0067)	−0.0109 (0.0065)

Education	Family Marketing	Sales activities through calls on customers	Tender calculations	Handling of orders	Advertising	Purchasing
Degree in engineering	−0.0054 (0.0023)	0.0308 (0.0562)	0.0552 (0.0351)	0.1413 (0.0702)	−0.0291 (0.0927)	0.0358 (0.0387)
Certificate in engineering I	−0.0008 (0.0025)	−0.0140 (0.0113)	0.0087 (0.0091)	−0.0080 (0.0151)	0.0039 (0.0240)	0.0043 (0.0125)
Certificate in engineering II	−0.0002 (0.0023)	0.0018 (0.0073)	−0.0068 (0.0050)	−0.0030 (0.0121)	−0.0295 (0.0250)	0.0179 (0.0092)
Degree in business & economics	0.0080 (0.0058)	−0.0764 (0.0322)	–	−0.0069 (0.0918)	−0.0215 (0.0282)	0.0250 (0.0302)
Certificate in commerce	0.0039 (0.0042)	0.0287 (0.0132)	−0.0816 (0.0697)	0.1138 (0.0213)	0.0125 (0.0221)	−0.0272 (0.0085)

	Sum of squares	d.f.
Total	491.5437	5,222
Explained	367.0096	81
Residual	124.5341	5,141

TABLE VIII f: P64, number of observations

Total number of observations: 2,429

Age	
–25	218
26–29	382
30–34	461
35–44	762
45–59	526
60–	80

Education	
Degree in engineering	45
Certificate in engineering I	212
Certificate in engineering II	258
Degree in business & economics	446
Certificate in commerce	1468

Job

Family	Level					
	2	3	4	5	6	7
Financial administration	127	210	286	171	548	211
Accounting, budget and cashier work	33	103	202	179		
Internal auditing		20	36	26		
Office rationalization		34	55	56		
Administrative rationalization		17	36	60	18	1

Interaction

Age	Education				
	Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce
–29	10	85	65	49	391
30–44	30	85	106	290	712
45–	5	42	87	107	365

Age	Level					
	2	3	4	5	6	7
–29	0	9	63	170	258	100
30–44	87	245	405	231	200	55
45–	73	130	147	91	108	57

Age	Family				
	Financial administration	Accounting, budget and cashier work	Internal auditing	Office rationalization	Administrative rationalization
–29	409	97	16	35	43
30–44	693	304	47	97	82
45–	451	116	19	13	7

Education	Level					
	2	3	4	5	6	7
Degree in engineering	0	18	18	6	3	
Certificate in engineering I	5	16	34	61	78	18
Certificate in engineering II	0	9	37	40	126	46
Degree in business & economics	107	158	135	33	10	3
Certificate in commerce	48	183	391	352	349	145

Education	Family				
	Financial administration	Accounting, budget and cashier work	Internal auditing	Office rationalization	Administrative rationalization
Degree in engineering	6	9	0	12	18
Certificate in engineering I	109	30	2	33	38
Certificate in engineering II	183	30	0	24	20
Degree in business & economics	208	147	34	36	21
Certificate in commerce	1047	301	45	40	35

TABLE VIIIf:P64,1

Intercept 7.8611 (0.0031)

R = 0.9122

Age

– 25	–0.2418	(0.0112)
26 – 29	–0.1154	(0.0079)
30 – 34	–0.0395	(0.0067)
35 – 44	0.0686	(0.0050)
45 – 59	0.1095	(0.0064)
60 –	0.0638	(0.0177)

Education

Degree in engineering	0.1786	(0.0248)
Certificate in engineering I	0.0088	(0.0110)
Certificate in engineering II	0.0166	(0.0100)
Degree in business & economics	0.0829	(0.0080)
Certificate in commerce	–0.0348	(0.0029)

Job

Family	Level					
	2	3	4	5	6	7
Financial administration	0.5082 (0.0162)	0.2958 (0.0117)	0.0763 (0.0101)	–0.0746 (0.0131)	–0.2449 (0.0071)	–0.3812 (0.0111)
Accounting, budget and cashier work	0.5924 (0.0295)	0.3173 (0.0166)	0.0901 (0.0110)	–0.1130 (0.0128)		
Internal auditing		0.3450 (0.0407)	0.1666 (0.0265)	–0.0847 (0.0354)		
Office rationalization		0.3093 (0.0400)	0.0565 (0.0292)	–0.0997 (0.0305)		
Administrative rationalization		0.3957 (0.0453)	0.2137 (0.0337)	–0.0201 (0.0282)	–0.1805 (0.0461)	

Interaction

Age	Education				
	Degree in engineering	Certificate in engineering I	Certificate in engineering II	Degree in business & economics	Certificate in commerce
– 29	0.0234 (0.0561)	0.0166 (0.0150)	0.0240 (0.0171)	0.0535 (0.0263)	–0.0149 (0.0048)
30 – 44	–0.0117 (0.0222)	0.0062 (0.0143)	0.0049 (0.0124)	0.0028 (0.0059)	–0.0021 (0.0030)
45 –	0.0232 (0.0750)	–0.0461 (0.0228)	–0.0239 (0.0147)	–0.0321 (0.0148)	0.0201 (0.0050)

Age	Level					
	2	3	4	5	6	7
– 29	–	–0.1536 (0.0540)	–0.0604 (0.0209)	–0.0150 (0.0104)	0.0067 (0.0074)	0.0601 (0.0118)
30 – 44	–0.0538 (0.0123)	–0.0152 (0.0058)	0.0037 (0.0042)	0.0213 (0.0078)	0.0117 (0.0093)	–0.0069 (0.0185)
45 –	0.0641 (0.0146)	0.0392 (0.0108)	0.0156 (0.0104)	–0.0262 (0.0148)	–0.0376 (0.0142)	–0.0988 (0.0180)

Age	Family				
	Financial administration	Accounting, budget and cashier work	Internal auditing	Office rationalization	Administrative rationalization
– 29	0.0052 (0.0063)	–0.0022 (0.0178)	0.0016 (0.0402)	–0.0442 (0.0280)	–0.0093 (0.0241)
30 – 44	0.0000 (0.0036)	0.0001 (0.0059)	–0.0166 (0.0152)	0.0163 (0.0104)	–0.0103 (0.0132)
45 –	–0.0048 (0.0039)	0.0016 (0.0129)	0.0396 (0.0331)	–0.0029 (0.0437)	0.1780 (0.0677)

Education	Level					
	2	3	4	5	6	7
Degree in engineering	–	−0.0313 (0.0355)	0.0314 (0.0297)	−0.0019 (0.0819)	–	–
Certificate in engineering I	−0.0696 (0.0702)	−0.0120 (0.0400)	0.0085 (0.0273)	0.0198 (0.0225)	−0.0033 (0.0198)	−0.03 (0.03
Certificate in engineering II	–	−0.0013 (0.0565)	0.0081 (0.0309)	0.0242 (0.0330)	−0.0060 (0.0142)	−0.01 (0.02
Degree in business & economics	0.0038 (0.0088)	−0.0128 (0.0083)	−0.0117 (0.0104)	0.0318 (0.0270)	0.2156 (0.0506)	–
Certificate in commerce	−0.0013 (0.0193)	0.0152 (0.0081)	0.0011 (0.0049)	−0.0091 (0.0055)	−0.0033 (0.0060)	0.00 (0.00

Education	Family				
	Financial administration	Accounting, budget and cashier work	Internal auditing	Office rationa-lization	Administrative rationalization
Degree in engineering	0.0168 (0.0659)	0.0022 (0.0505)	–	0.0247 (0.0436)	−0.0232 (0.0345)
Certificate in engineering I	0.0161 (0.0154)	−0.0156 (0.0303)	–	0.0008 (0.0231)	−0.0344 (0.0293)
Certificate in engineering II	0.0094 (0.0114)	−0.0399 (0.0376)	–	−0.0224 (0.0321)	0.0003 (0.0444)
Degree in business & economics	−0.0187 (0.0097)	0.0066 (0.0110)	0.0374 (0.0233)	0.0481 (0.0372)	−0.0034 (0.0378)
Certificate in commerce	0.0003 (0.0023)	0.0023 (0.0062)	−0.0283 (0.0176)	0.0068 (0.0343)	−0.0000 (0.0289)

	Sum of squares	d.f.
Total	327.9332	2,428
Explained	272.8897	81
Residual	55.0735	2,347

TABLE VIIIg:H68,3 x)

Intercept 8.2183 (0.0049) R = 0.7578

Age		
– 25	–0.4550	(0.0065)
26 – 29	–0.2190	(0.0053)
30 – 34	–0.0071	(0.0050)
35 – 44	0.1967	(0.0042)
45 – 59	0.2858	(0.0045)
60 –	0.1986	(0.0102)

Cost of living area		
3	–0.0465	(0.0031)
4	–0.0014	(0.0039)
5	0.0479	(0.0036)

Education		
Degree in engineering	0.2077	(0.0069)
Certificate in engineering I	–0.0731	(0.0043)
Certificate in engineering II	–0.1495	(0.0040)
Degree in business & economics	0.1575	(0.0068)
Certificate in commerce	–0.1426	(0.0044)

Industry		
Mining	0.0227	(0.0295)
Metal and engineering industry		
Iron and steel works, metal plants	0.0141	(0.0090)
Manufacture of hardware	0.0008	(0.0109)
Engineering works	–0.0153	(0.0056)
Repair works	–0.0006	(0.0149)
Shipyards	–0.0700	(0.0154)
Manufacture of electrical equipment	–0.0549	(0.0066)
Other metal industry	–0.0087	(0.0158)
Quarrying: stone, clay and glass products	0.0236	(0.0109)
Wood industry	0.0171	(0.0149)
Manufacture of pulp, paper and paper products	0.0431	(0.0114)
Printing and allied industries	–0.0061	(0.0222)
Food manufacturing industries	–0.0028	(0.0142)
Beverage and tobacco industries	–0.0063	(0.0360)
Textile industry	0.0549	(0.0145)
Leather, furs and rubber industries	–0.0253	(0.0163)
Chemical industry	0.0377	(0.0094)
Building and construction	–0.0241	(0.0071)

	Sum of squares	d.f.
Total	1469.0762	11,819
Explained	843.6416	28
Residual	625.4347	11,791

x) For number of observations see table VIIIa:H68

TABLE IX:H68, number of observations

Total number of observations: 50,121

Age

– 19	28
20 – 21	245
22 – 23	1,431
24 – 25	3,185
26 – 27	3,911
28 – 29	3,863
30 – 31	3,805
32 – 34	4,919
35 – 39	7,237
40 – 44	7,211
45 – 49	6,422
50 – 59	6,485
60 –	1,379

Cost of living area

3	29,299
4	8,507
5	12,315

Education

Degree in engineering	5,660
Certificate in engineering I	16,121
Certificate in engineering II	22,527
Degree in business & economics	1,010
Certificate in commerce	3,524
Degree in social work public administration	62
Degree in science	423
Degree in law or in the social sciences	252
Other university degrees	542

Job family

0.	Administration	1,074
1.	Production supervision of control	9,649
2.	Research and development	6,972
3.	Design	12,303
4.	Other technical work	7,297
5.	Humanistic work	172
6.	Education	187
7.	General service and health	102
8.	Commercial	8,712
9.	Accounting and general office work	3,653

Job level

2	1,707
3	6,083
4	13,459
5	18,585
6	8,634
7	1,596
8	57

Industry

Mining	678
Metal and engineering industry	
Iron and steel works, metal plants	4,023
Manufacture of hardware	1,527
Engineering works	14,939
Repair works	377
Shipyards	1,590
Manufacture of electrical equipment	9,226
Other metal industry	592
Quarrying: stone, clay ang glass products	1,573
Wood industry	701
Manufacture of pulp, paper and paper products	1,986
Printing and allied industries	253
Food manufacturing industries	897
Beverage and tobacco industries	110
Textile industry	717
Leather, furs and rubber industry	590
Chemical industry	3,178
Building and construction	7,164

TABLE IX:H68,1

Intercept 8.1362 (0.0006)

R = 0.9229

Age

– 19	–0.5184	(0.0257)
20 – 21	–0.3929	(0.0089)
22 – 23	–0.3030	(0.0038)
24 – 25	–0.2300	(0.0025)
26 – 27	–0.1669	(0.0022)
28 – 29	–0.1081	(0.0022)
30 – 31	–0.0597	(0.0021)
32 – 34	–0.0117	(0.0019)
35 – 39	0.0441	(0.0015)
40 – 44	0.0910	(0.0015)
45 – 49	0.1134	(0.0017)
50 – 59	0.1252	(0.0016)
60 –	0.0847	(0.0036)

Cost of living area

3	–0.0282	(0.0005)
4	0.0033	(0.0014)
5	0.0648	(0.0012)

Education

Degree in engineering	0.1497	(0.0019)
Certificate in engineering I	–0.0022	(0.0009)
Certificate in engineering II	–0.0335	(0.0008)
Degree in business & economics	0.0488	(0.0046)
Certificate in commerce	–0.0508	(0.0028)
Degree in social work public administration	–0.1151	(0.0175)
Degree in science	0.0710	(0.0067)
Degree in law or in the social sciences	0.0160	(0.0088)
Other degrees	0.0835	(0.0061)

Job family

0.	Administration	0.0651	(0.0044)
1.	Production supervision of control	0.0338	(0.0014)
2.	Research and development	–0.0179	(0.0016)
3.	Design	–0.0290	(0.0011)
4.	Other technical work	–0.0327	(0.0015)
5.	Humanistic work	–0.0743	(0.0104)
6.	Education	–0.0489	(0.0099)
7.	General service and health	0.1613	(0.0138)
8.	Commercial	0.0413	(0.0014)
9.	Accounting and general office work	–0.0081	(0.0027)

Job level

2	0.5345	(0.0035)
3	0.3249	(0.0018)
4	0.1064	(0.0011)
5	–0.0859	(0.0008)
6	–0.2471	(0.0016)
7	–0.3561	(0.0037)
8	–0.4015	(0.0181)

Industry

Mining	0.0090	(0.0052)
Metal and engineering industry		
Iron and steel works, metal plants	0.0116	(0.0021)
Manufacture of hardware	0.0140	(0.0034)
Engineering works	–0.0130	(0.0010)
Repair works	0.0228	(0.0070)
Shipyards	–0.0126	(0.0034)
Manufacture of electrical equipment	–0.0082	(0.0014)
Other metal industry	–0.0005	(0.0055)
Quarrying: stone, clay and glass products	0.0252	(0.0034)
Wood industry	0.0128	(0.0051)
Manufacture of pulp, paper and paper products	0.0184	(0.0030)
Printing and allied industries	0.0280	(0.0086)
Food manufacturing industries	–0.0117	(0.0045)
Beverage and tobacco industries	0.0355	(0.0130)
Textile industry	0.0046	(0.0051)
Leather, furs and rubber industries	–0.0095	(0.0056)
Chemical industry	0.0127	(0.0024)
Building and construction	0.0117	(0.0017)

	Sum of squares	d.f.
Total		50,120
Explained		54
Residual		50,066

APPENDIX B

NECESSARY AND SUFFICIENT CONDITIONS FOR UNIQUE SOLUTIONS IN CONSTRAINED LINEAR MODELS NOT OF FULL RANK

Suppose Z is an nxc matrix of rank r, $r < c$, H is a txc matrix, τ is a column vector with c elements and g a column vector with n elements.

Proposition B:1

For every g in the column space of Z there is a unique solution τ to the equation equation system

$$\begin{cases} Z\tau = g; & \text{(B:1a)} \\ H\tau = 0; & \text{(B:1b)} \end{cases}$$

if and only if

i) rank $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix} = c$;

ii) no linear combination of the rows of H is a linear combination of the rows of Z except the null vector.

Proof

We first prove the *necessity* of the proposition. The equation system (B:1a–b) is rewritten

$$\begin{Bmatrix} Z \\ \hline H \end{Bmatrix} \tau = \begin{Bmatrix} g \\ \hline 0 \end{Bmatrix}; \tag{B:2}$$

Suppose there is a unique solution to this equation system for every g in the column space of Z. It then follows that the columns of the matrix $\begin{Bmatrix} Z \\ \hline H \end{Bmatrix}$ are linearly independent which implies the first condition of the proposition.

Premultiplication of (B:2) by $\{Z' \mid H'\}$ gives

$$(Z'Z + H'H)\tau = Z'g, \tag{B:3}$$

As the rank of $\left\{\frac{Z}{H}\right\}$ is c the inverse of $(Z'Z + H'H)$ exists. The unique solution to (B:3) and (B:2) is

$$\tau = (Z'Z + H'H)^{-1}Z'g; \qquad \text{(B:4)}$$

The general solution to (B:1a) is

$$\tau = Z^{-}g + (I-Z^{-}Z)k; \qquad \text{(B:5)}$$

where k is an arbitrary c×1 vector and Z^{-} a g_1-inverse to Z, (see note p. 140).

As (B:4) is a solution to (B:2) it is also one of the solutions (B:5) and for every g there is a Z^{-} and a k which satisfy

$$Z^{-}g + (I-Z^{-}Z)k = (Z'Z+H'H)^{-1}Z'g \qquad \text{(B:6)}$$

Premultiplication by Z gives

$$ZZ^{-}g = Z(Z'Z+H'H)^{-1}Z'g; \qquad \text{(B:7)}$$

$(Z'Z+H'H)^{-1}Z'$ is thus a g_1-inverse to Z, i.e.

$$Z^{-} = (Z'Z+H'H)^{-1}Z'; \qquad \text{(B:8)}$$

By definition it is true that

$$(Z'Z+H'H)^{-1}(Z'Z+H'H) = I; \qquad \text{(B:9a)}$$

and

$$(Z'Z+H'H)^{-1}Z'Z + (Z'Z+H'H)^{-1}H'H = I; \qquad \text{(B:9b)}$$

Premultiplication by Z gives

$$Z(Z'Z+H'H)^{-1}Z'Z + Z(Z'Z+H'H)^{-1}H'H = Z; \qquad \text{(B:10a)}$$

If (B:8) is used (B10a) can be rewritten as

$$Z + Z(Z'Z+H'H)^{-1}H'H = Z; \qquad \text{(10b)}$$

It follows that

$$Z(Z'Z+H'H)^{-1}H'H = 0; \tag{B:11}$$

If (B:9b) is premultiplied by H'H, (B:11) by Z' and the result is added side by side, the following expression is obtained

$$H'HZ^{-}Z + (Z'Z+H'H)(Z'Z+H'H)^{-1}H'H = H'H ; \tag{B:12}$$

which implies

$$H'HZ^{-}Z = 0; \tag{B:13}$$

After premultiplication by H'H (B:9b) can now be reduced to the following expression if (B:8) and (B:13) is used

$$H'H(Z'Z+H'H)^{-1}H'H = H'H; \tag{B:14}$$

i.e. $(Z'Z+H'H)^{-1}$ is a g_1-inverse to H'H. However, $(Z'Z+H'H)^{-1}$ is a symmetric matrix of full rank which satisfies the conditions of a »Moore-Penrose inverse».★ We write

$$(H'H)^{g} = (Z'Z+H'H)^{-1} ; \tag{B:15}$$

For a Moore-Penrose inverse it is true that

$$(H'H)^{g} = H^{g}(H^{g})'; \tag{B:16}$$

where H^g is the Moore-Penrose inverse to H (see for instance Pringle & Rayner [1971] p. 6). (B:11) can now be written

$$Z(H'H)^{g}H'H = 0; \tag{B:17a}$$

or

$$ZH^{g}(H^{g})'H'H = 0; \tag{B:17b}$$

★ A^g is a »Moore-Penrose inverse» to A if the following four conditions are satisfied:
a) $AA^gA = A$; b) $A^gAA^g = A^g$; c) $(AA^g)' = AA^g$; d) $(A^gA)' = A^gA$.

but $(H^g)'H'H = H$ and thus

$$ZH^gH = 0; \qquad \text{(B:17c)}$$

For any vectors u and v such that $u'Z + v'H = 0$ it is true that $u'Z = v'H = 0$, because from (B:17c) and the definition of a Moore-Penrose inverse we obtain

$$v'H = u'ZH^gH + v'HH^gH = (u'Z+v'H)H^gH = 0; \qquad \text{(B:18)}$$

i.e. condition ii) is true.

It has now been proved that the two conditions i) and ii) are necessary. It remains to show that they are also *sufficient.*

Suppose H is chosen such that i) and ii) are true. From the first condition it follows that the columns of $\left\{\begin{matrix} Z \\ \hline H \end{matrix}\right\}$ are linearly independent. If a solution exists for every g in the column space of Z it is thus a unique solution. From i) it also follows that there is an inverse to the matrix $(Z'Z+H'H)$. There is thus a unique solution (B:4) to (B:3). To prove that this solution is also a solution of (B:2) we observe that (B:10a) can be written as

$$\left\{ Z(Z'Z+H'H)^{-1}Z' - I\right\} Z + \left\{ Z(Z'Z+H'H)^{-1}H\right\} H = 0; \qquad \text{(B:19)}$$

Because of the second condition ii) it follows that (B:11) is true. After insertion of (B:11) into (B:10a) we find that $(Z'Z+H'H)^{-1}Z$, is a g_1-inverse to Z which shows that (B:4) is one of the solutions (B:5).

After premultiplication by H and rearrangement of the terms (B:9b) can be written

$$\left\{ H(Z'Z+H'H)^{-1}Z'\right\} Z + \left\{ H(Z'Z+H'H)^{-1}H' - I\right\} H = 0; \qquad \text{(B:20)}$$

Because of condition ii) it follows that

$$H(Z'Z+H'H)^{-1}Z'Z = 0; \qquad \text{(B:21)}$$

To see that the solution (B:4) satisfies the constraints (B:1b) we only have to observe that

$$H\tau = H(Z'Z+H'H)^{-1}Z'g = H(Z'Z+H'H)^{-1}Z'Z\tau = 0; \qquad \text{(B:22)}$$

This *ends the proof* of sufficiency and the whole proposition B:1 is proved.

Three corollaries follow.

Corollary B:1. If the constraints (B:1b) are chosen such that the conditions i) and ii) are satisfied, the unique solution to (B:2) is

$$\tau = (Z'Z+H'H)^{-1}Z'g; \tag{B:23}$$

This corollary follows from the proof above.

Corollary B:2. Matrices Z and H which satisfy the conditions of proposition B:1 are complementary.

To prove this corollary we first remember that rank Z is r and that Z and H both have c columns. Suppose in addition that rank H is s. The rows of Z then span an r-dimensional space and the rows of H an s-dimensional space. Because of condition ii) the rows of the augmented matrix $\left\{\frac{Z}{H}\right\}$ span an r+s-dimensional space. From condition i) it then follows that r+s=c and that Z and H are complementary.

Corollary B:3. The condition i) and ii) in proposition B:1 are equivalent to the following conditions

i) $\text{rank}\left\{\frac{Z}{H}\right\} = c;$

iii) $\text{rank } Z + \text{rank } H = \text{rank}\left\{\frac{Z}{H}\right\};$

As we have already shown that Z and H are complementary, the necessity of corollary B:3 is proved. If there were vectors u and v, such that $u'Z+v'H=0$, but $u'Z\neq0$ and $v'H\neq0$, then rank $\left\{\frac{Z}{H}\right\}$ is less than the sum of rank Z and rank H, which is contrary to condition iii). iii) is thus a sufficient condition for the truth of ii). The sufficiency of the whole corollary is then proved.

APPENDIX C

SOME PROPERTIES OF LEAST SQUARES REGRESSION WITH INCOMPLETE MODELS NOT OF FULL RANK

Proposition C:1

Given the model (5:94) and the support models (5:95) and (5:96), if $H_{11} = H_{21} = H_1$; $H_{12} = H_2$ and $H_{23} = H_3$ then the least squares estimate of an element of τ_1 is a weighted sum of the elements of the least squares estimate $\hat{\tau}_{11}$ and $\hat{\tau}_{21}$ and the least squares estimate of τ_2 (τ_3) equals the least squares estimate $\hat{\tau}_{12}$ ($\hat{\tau}_{23}$) plus a weighted sum of the elements of $\hat{\tau}_{11}$ and $\hat{\tau}_{21}$.

Proof

The normal equations to the model (5:94) are

$$\left\{\begin{array}{c|c|c} Z'_{11}Z_{11} + Z'_{21}Z_{21} + H'_1H_1 & Z'_{11}Z_{12} + H'_1H_2 & Z'_{21}Z_{23} + H'_1H_3 \\ \hline Z'_{12}Z_{11} + H'_2H_1 & Z'_{12}Z_{12} + H'_2H_2 & H'_2H_3 \\ \hline Z'_{23}Z_{21} + H'_3H_1 & H'_3H_2 & Z'_{23}Z_{23} + H'_3H_3 \end{array}\right\} \left\{\begin{array}{c} \hat{\tau}_1 \\ \hline \hat{\tau}_2 \\ \hline \hat{\tau}_3 \end{array}\right\} =$$

$$= \left\{\begin{array}{c} Z'_{11}\ell_1 + Z'_{21}\ell_2 \\ \hline Z'_{12}\ell_1 \\ \hline Z'_{23}\ell_2 \end{array}\right\} ; \qquad \text{(C:1)}$$

and the normal equations to the two support models are respectively

$$\left\{\begin{array}{c|c} Z'_{11}Z_{11} + H'_{11}H_{11} & Z'_{11}Z_{12} + H'_{11}H_{12} \\ \hline Z'_{12}Z_{11} + H'_{12}H_{11} & Z'_{12}Z_{12} + H'_{12}H_{12} \end{array}\right\} \left\{\begin{array}{c} \hat{\tau}_{11} \\ \hline \hat{\tau}_{12} \end{array}\right\} = \left\{\begin{array}{c} Z'_{11}\ell_1 \\ \hline Z'_{12}\ell_1 \end{array}\right\} ; \qquad \text{(C:2)}$$

and

$$\left\{\begin{array}{c|c} Z'_{21}Z_{21} + H'_{21}H_{21} & Z'_{21}Z_{23} + H'_{21}H_{23} \\ \hline Z'_{23}Z_{21} + H'_{23}H_{21} & Z'_{23}Z_{23} + H'_{23}H_{23} \end{array}\right\} \left\{\begin{array}{c} \hat{\tau}_{21} \\ \hline \hat{\tau}_{23} \end{array}\right\} = \left\{\begin{array}{c} Z'_{21}\ell_2 \\ \hline Z'_{23}\ell_2 \end{array}\right\} ; \qquad \text{(C:3)}$$

$$\hat{\tau}_{1J} = \frac{\begin{Vmatrix} (Z'_{11}Z_{11}+H'_{11}H_{11})\hat{\tau}_{11} + (Z'_{11}Z_{12}+H'_{11}H_{12})\hat{\tau}_{12} + (Z'_{21}Z_{21}+H'_{21}H_{21})\hat{\tau}_{21} + (Z'_{21}Z_{23}+H'_{21}H_{23})\hat{\tau}_{23} & (Z'_{11}Z_{11}+Z'_{21}Z_{21}+H'_{1}H_{1})_{-J} & Z'_{11}Z_{12}+H'_{1}H_{2} & Z'_{21}Z_{23}+H'_{1}H_{3} \\ (Z'_{12}Z_{11}+H'_{12}H_{11})\hat{\tau}_{11} + (Z'_{12}Z_{12}+H'_{12}H_{12})\hat{\tau}_{12} & (Z'_{12}Z_{11}+H'_{2}H_{1})_{-J} & Z'_{12}Z_{12}+H'_{2}H_{2} & H'_{2}H_{3} \\ (Z'_{23}Z_{21}+H'_{21}H_{21})\hat{\tau}_{21} + (Z'_{23}Z_{23}+H'_{23}H_{23})\hat{\tau}_{23} & (Z'_{23}Z_{21}+H'_{3}H_{1})_{-J} & H'_{3}H_{2} & Z'_{23}Z_{23}+H'_{3}H_{3} \end{Vmatrix}}{\Vert X'X+H'H \Vert} (-1)^{J-1}; \quad \text{(C:4)}$$

$$\hat{\tau}_{1J} = \frac{\begin{Vmatrix} (Z'_{11}Z_{11}+H'_{11}H_{11})\hat{\tau}_{11} & \\ (Z'_{12}Z_{11}+H'_{12}H_{11})\hat{\tau}_{11} & S \\ 0 & \end{Vmatrix}}{\Vert X'X+H'H \Vert}(-1)^{J-1} + \frac{\begin{Vmatrix} (Z'_{11}Z_{12}+H'_{12}H_{12})\hat{\tau}_{12} & \\ (Z'_{12}Z_{12}+H'_{12}H_{12})\hat{\tau}_{12} & S \\ 0 & \end{Vmatrix}}{\Vert X'X+H'H \Vert}(-1)^{J-1} + \frac{\begin{Vmatrix} (Z'_{21}Z_{21}+H'_{21}H_{21})\hat{\tau}_{21} & \\ 0 & S \\ (Z'_{23}Z_{21}+H'_{23}H_{21})\hat{\tau}_{21} & \end{Vmatrix}}{\Vert X'X+H'H \Vert}(-1)^{J-1} + \frac{\begin{Vmatrix} (Z'_{21}Z_{23}+H'_{21}H_{23})\hat{\tau}_{23} & \\ 0 & S \\ (Z'_{23}Z_{23}+H'_{23}H_{23})\hat{\tau}_{23} & \end{Vmatrix}}{\Vert X'X+H'H \Vert}(-1)^{J-1}; \quad \text{(C:5)}$$

$$\begin{Vmatrix} (Z'_{11}Z_{12}+H'_{11}H_{12})\hat{\tau}_{12} & \\ (Z'_{12}Z_{12}+H'_{12}H_{12})\hat{\tau}_{12} & S \\ 0 & \end{Vmatrix} = \begin{Vmatrix} (Z'_{11}Z_{12}+H'_{11}H_{12})\hat{\tau}_{12} & \\ (Z'_{12}Z_{12}+H'_{12}H_{12})\hat{\tau}_{12} & S \\ H'_{23}H_{12}\,\hat{\tau}_{12} & \end{Vmatrix} = \sum_{k=1}^{k_2} \begin{Vmatrix} (Z'_{11}Z_{12}+H'_{11}H_{12})_{+k} & \\ (Z'_{12}Z_{12}+H'_{12}H_{12})_{+k} & S \\ (H'_{23}H_{12})_{+k} & \end{Vmatrix} \hat{\tau}_{12k}; \quad \text{(C:6)}$$

$$\hat{\tau}_{1J} = \sum_{k=1}^{k_1} \frac{\begin{Vmatrix} (Z'_{11}Z_{11}+H'_{11}H_{11})_{+k} & \\ (Z'_{12}Z_{11}+H'_{12}H_{11})_{+k} & S \\ 0 & \end{Vmatrix}}{\Vert X'X+H'H \Vert} \hat{\tau}_{11k}(-1)^{J-1} + \sum_{k=1}^{k_1} \frac{\begin{Vmatrix} (Z'_{21}Z_{21})_{+k} & \\ 0 & S \\ (Z'_{23}Z_{21}+H'_{23}H_{21})_{+k} & \end{Vmatrix}}{\Vert X'X+H'H \Vert} \hat{\tau}_{21k}(-1)^{J-1}; \quad \text{(C:7)}$$

(C:2) and (C:3) are now substituted block by block into the vector to the right of the equality sign in (C:1) and the resulting normal equations are solved by using Cramer's rule.

The following notations are introduced:

$\hat{\tau}_{iJ}$ = the J-th element of the vector $\hat{\tau}_i$, i=1,2 and 3.

$\hat{\tau}_{11k}$ = the k-th element of the vector $\hat{\tau}_{11}$.

Analogous notations are used for $\hat{\tau}_{12}$, $\hat{\tau}_{21}$ and $\hat{\tau}_{23}$.

$(A)_{+k}$ = the k-th column of the matrix A

$(A)_{-k}$ = a matrix of all columns of A except the k-th

The expression for $\hat{\tau}$ is obtained from Cramer's rule as given in (C:4). Following elementary rules of matrix algebra the determinants in the numerator can be written as a sum of four determinants, each with the last three blocks of the numerator determinant in (C:4) in common. To simplify the notations the matrices formed by these three blocks are denoted by S. The new expression obtained for $\hat{\tau}_1$ is (C:5). Before the same elementary operation is repeated on the determinant in the numerator of the second term of (C:5), we note that $H'_{23}H'_{12}\hat{\tau}_{12} = 0$ because of the constraint imposed on the least squares estimates. The result is spelt out in expression (C:6). If it is true that $H_{11} = H_1$; $H_{12} = H_2$; and $H_{23} = H_3$; then it is easy to see, recalling the definition of S, that each of the determinants in (C:6) vanishes because two columns in each determinant are identically the same. The second term in the sum (C:5) thus vanishes and in the same way it can be proved that the fourth term vanishes. The expression for $\hat{\tau}_{iJ}$ can now be simplified to (C:7). To obtain $\hat{\tau}_{1j}$ in this form the same operation as above has been used on the determinants and also the homogeneous constraint on $\hat{\tau}_{21}$ which implies that

$$H'_{21}H_{21}\hat{\tau}_{21} = 0; \qquad \text{(C:8)}$$

Alternatively the constraint on τ_{11} could have been used.

The two terms with k=J are separated from the remainder and (C:7) is re-written as

$$\hat{\tau}_{1J} = \omega_{11J}\hat{\tau}_{11J} + \omega_{21J}\hat{\tau}_{21J} + \sum_{\substack{k \\ k \neq J}} (\omega_{11k}\hat{\tau}_{11k} + \omega_{21k}\hat{\tau}_{21k}), \qquad \text{(C:9)}$$

where ω_{11k} and ω_{21k} are weights equal to the ratios of determinants in (C:7).

If again

$$H_{11} = H_{21} = H_1 ; H_{12} = H_2 \text{ and } H_{23} = H_3 ; \tag{C:10}$$

we find that

$$\omega_{11J} + \omega_{21J} = 1; \qquad J=1,...,k_1 \tag{C:11}$$

$$\omega_{11k} + \omega_{21k} = 0; \qquad k \neq J \tag{C:12}$$

Following the same procedure as above it can be shown that if (C:10) is true then

$$\hat{\tau}_{2J} = \hat{\tau}_{12J} + \sum_{k=1}^{k_1} (\omega^*_{11k} \tau_{11k} + \omega^*_{21k} \hat{\tau}_{21k}); \qquad J=1,...,k_2 \tag{C:13}$$

$$\hat{\tau}_{3J} = \hat{\tau}_{23J} + \sum_{k=1}^{k_1} (\omega^{**}_{11k} \hat{\tau}_{11k} + \omega^{**}_{21k} \hat{\tau}_{21k}); \qquad J=1,...,k_3 \tag{C:14}$$

ω^*_{ijk} and ω^{**}_{ijk} are weights built up from determinants of the same principal structure as those in expression (C:5). By pairs these weights add up to zero.

$$\omega_{11k} + \omega^*_{21k} = 0; \qquad k=1,...,k_1 \tag{C:15}$$

$$\omega^{**}_{11k} + \omega^{**}_{21k} = 0; \qquad k=1,...,k_1 \tag{C:16}$$

End of proof

REFERENCES

Andrews, F.M., Morgan, J.N., Sonquist, J.A., 1969, *Multiple Classification Analysis.* A Report on a Computer Program for Multiple regression Using Categorical Predictors. Survey Research Center, Institute for Social Research. The University of Michigan. Ann Arbor.

Bailey, D. & Schotta, C., 1969, The Rate of Return to Undergraduate Education. *American Statistical Association,* 1969 Proceedings of the Business and Economic Section, pp. 460 f.

Becker, G.S., 1962, Investment in Human Capital: A Theoretical Analysis. *The Journal of Political Economy.* Vol. LXX, Suppl, Oct. 1962, No. 5, Part 2, pp. 9–49.

– , 1964, *Human Capital.* Columbia University Press, New York.

Ben-Porath, Y., 1966, Lifetime Income and Economic Growth: Comment. *The American Economic Review.* Vol. LVI. Sept. 1966, No. 4, Part 1, pp. 869–872.

Blaug, M., Peston, M.H. & Ziderman, A., 1967, *The Utilization of Educated Manpower in Industry.* Edinburgh and London.

Cedheim, B. & Klevmarken, A., 1969, *COMRAN. A Fortran Program for Constrained Multiple Regression Analysis.* Research Report from the Institute of Statistics. The University of Stockholm. Stockholm.

Chipman, J.S., 1964. On Least Squares with Insufficient Observations. *Journal of the American Statistical Association.* Vol. 59. Dec. 1964, No. 308. pp. 1078–1111.

David, M., 1969, Time-Series Versus Cross Section Lifetime Earnings Patterns in Different Occupational Groups. *American Statistical Association.* 1969 Proceedings of the Business and Economic Section, pp. 664 f.

Doll, R., 1971, The Age Distribution of Cancer: Implications for Models of Carcinogenesis (with Discussion). *Journal of the Royal Statistical Society.* Series A (General). Vol. 134. Part 2, pp. 133–166.

Edgren, G., Faxén, K.-O., Odhner, C.-E., 1968, *Lönebildning och samhällsekonomi.* (Wage Determination and the National Economy, Report of a Group of Experts Appointed by SAF, LO and TCO.)

Eriksson, I., 1970, *Ålder och inkomst.* Låginkomstutredningen . (Age and Income. Draft of chapter 9 in the Investigation of the Living Conditions of the Swedish People.) Stockholm

Eriksson, J.-E., 1968, Facklig verksamhet på den enskilda sektorn. (Trade Union Activity in the Private Sector.) *SACO-tidningen* No.7, 1968.

Faragó, G., 1968, Yrkesbyte i samband med anställningsbyte. (Change of Occupation in Connection with Change of Employment.) *Reports from National Labour Market Board.* No. 12, 1968.

Fase, M.M.G., 1969, *An Econometric Model of Age-Income Profiles. A Statistical Analysis of Dutch Income Data 1958–1967.* University of Rotterdam. Rotterdam.

Gadd, A. & Wold, H.O.A., 1964, Econometric Model Building, *Essays on the Causal Chain Approach.* (Ed. H. Wold) Amsterdam.

Hill, T.P. , 1959, An Analysis of the Distribution of Wages and Salaries in Great Britain. *Econometrica.* Vol. 27. July 1959, No. 3, pp.355–381.

Hofsten, E., 1971, Birth Variations in Populations which Practise Family Planning. *Population Studies.* Vol. XXV. No. 2, pp. 315–326.

Holm, M., 1970, Inkomstskillnader 1966 för helårs- och heltidsanställda män och kvinnor. (Income Differences 1966 for Men and Women Employed Full Time and for the Whole Year.) Chapter 5 and appendices 1–2 in Incomes of the Swedish People. *SOU* 1970:34.

Holmberg, P,. 1969, Våra okända lokala förhandlingar. (Our Unknown Local Negotiations.) *TCO-tidningen.* No. 7, 1969.

Klevmarken, A., 1967, *En statistisk deskription av industritjänstemännens utbildningssammansättning 1956–1964.* (A Statistical Description of the Educational Composition Among Salary Earners in Swedish Industry 1956–1964.) Discussion Paper at the Institute of Statistics, University of Stockholm and the Forecasting Institute, National Central Bureau of Statistics. Stockholm.

– , 1968a, *Lönestrukturen bland industriens tjänstemän 1957–1964.* (Salary Structure Among Salaried Employees in Industry 1957–1964.) Research Report from Institute of Statistics. The University of Stockholm. Stockholm.

– , 1968b, Aggregation over Sets of Observations. Some Properties of the Method of Least Squares when the Normal Equations are Underdetermined. *Statistisk Tidskrift.* Vol. 6. No. 5, 1968, pp. 372–380.

– , 1968c, Civilingenjörernas lönebildning. (Salary Determination of Graduate Engineers.) *CF-tidskriften.* No. 9, 1968. (The Swedish Association of Graduate Engineers.) Stockholm.

– , 1968d, *Modeller beskrivande ålders–utbildnings–inkomst profiler, med speciell tillämpning på industritjänstemän i Sverige.* (Models for Age–education–earnings Profiles with Special Application to Salaried Employees in Sweden.) Discussion paper at the Institute of Statistics, University of Stockholm and the Forecasting Institute, National Central Bureau of Statistics. Stockholm.

– , 1970, *Konsumtionsfunktioner uppskattade från tvärsnitt och från tidsserier. Konsumtion–inkomst–åldersprofiler.* (Consumption Function Estimated from Cross Sections and from Time Series. Consumption, Income and Age Profiles.) Research Report from Institute of Statistics. University of Stockholm. Stockholm.

– , 1971, Den svåra konsten att räkna inkomst- och löneökningar. (The Difficult Art of Calculating Income and Wage Increases.) *Industriförbundets Tidskrift* No. 4. 1971.

Lind, E., 1963, *Salary Differentiation in Swedish Industry.* Publications issued by the Swedish Employers' Confederation No. 17. Stockholm.

Lundquist, S., 1970, *On Least Squares Computations.* Research Report from Institute of Statistics. University of Stockholm. Stockholm.

Lydall, H., 1955, The Life Cycle in Income, Saving, and Asset Ownership, *Econometrica* Vol. 23. April 1955, No. 2, pp. 131–150.

– , 1968, *The Structure of Earnings.* Oxford University Press. Oxford.

Malinvaud, E., 1966, *Statistical Methods of Econometrics.* Amsterdam.

Malmquist, S., 1948, *A Statistical Analysis of the Demand for Liquor in Sweden.* A Study of the Demand for a Rationed Commodity. Uppsala.

Melichar, E., 1965, Least Squares Analysis of Economic Survey Data. *American Statistical Association.* 1965 Proceedings of the Business and Economic Statistics Section, pp. 373–385.

Merrett, S., 1966, The Rate of Return to Education: A Critique. *Oxford Economic Papers.* Vol. 18. Nov. 1966, No. 3, pp. 289–303.

Miller, H.P., 1965, Lifetime Income and Economic Growth. *The American Economic Review.*, Vol. LV. Sept. 1965, No. 4, pp. 834–844.

Mincer, J., 1958, Investment in Human Capital and Personal Income Distribution. *The Journal of Political Economy.* Vol. LXVI. Aug. 1958, No. 4, pp. 281–302.

– , 1970, The Distribution of Labor Incomes: A Survey. *Journal of Economic Literature.* Vol. VIII. March 1970, No. 1, pp. 1–26.

Morgan, J., 1962, The Anatomy of Income Distribution. *The Review of Economics and Statistics.* Vol. XLIV. Aug. 1962, No. 3, pp. 270–283.

Morgan, J.N. & Sonquist, J.A., 1963, Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association.* Vol. 58. June 1963, No. 302, pp. 415–434.

Plackett, R.L., 1950, Some Theorems in Least Squares. *Biometrica.* Vol. 37. June 1950, Parts 1 and 2, pp. 149–157.

– , 1960, *Principles of Regression Analysis.* Oxford.

Positional Classification System. Salaried Employees, 1968, The Swedish Employers' Confederation, The Swedish Union of Clerical and Technical Employees in Industry, The Swedish Foremen's and Supervisor's Union and The Swedish Union of Commercial Employees.

Pringle, R.M. & Rayner, A.A., 1971, *Generalized Inverse Matrices with Applications to Statistics.* Griffin's Statistical Monographs & Courses. London.

Rundblad, B.G., 1964, *Arbetskraftens rörlighet. En studie av en lokal arbetsmarknad.* (The Mobility of Labour. A Study of a Local Labour Market.) Industriens Utredningsinstitut. Stockholm.

SACO, 1968, *Skenbar och verklig inkomst.* (Apparent and Real Income.) The Swedish Confederation of Professional Associations. Stockholm.

SAF, 1969, *Utjämnade lönekurvor.* (Smoothed Earnings Curves.) The Swedish Employers' Confederation. Stockholm.

SAF, yearly, *Löner för tjänstemän.* (Earnings of Salaried Employees.) The Swedish Employers' Confederation. Stockholm.

SAF–SIF, 1968, The Swedish Employers' Confederation – The Swedish Union of Clerical and Technical Employees in Industry. *Lönepolitik.* (Salary Policies.) Report of the Salary Policy Committee. Stockholm.

Scheffé, H., 1959, *The Analysis of Variance.* New York and London.

Searle, S.R., 1971, *Linear Models.* New York, London, Sydney, Toronto.

Seber, G.A.F., 1966, *The Linear Hypothesis: A General Theory.* London.

Siegfrid, J.J., 1971, Rate of Return to Ph.D. in Economics. *Industrial and Labor Relations Review.* Vol. 24. April 1971, No. 3, pp. 420–431.

SIF–SALF–CF, 1970, *Löneavtal 1970–74.* (Salary Agreements 1970–74.) Stockholm.

SOU, 1970:34, see Holm [1970].

Suits, D.B., 1957. Use of Dummy Variables in Regression Equations. *Journal of the American Statistical Association.* Vol. 52. Dec. 1957, No. 280, pp. 548–551.

TCO, 1966, The Central Organization of Salaried Employees in Sweden, *Job Classification and Collective Bargaining.* Papers and Discussions from an International Seminar at Gällöfsta, Sweden, June 6–10, 1966.

Tobin, J., 1967, Life Cycle Saving and Balanced Growth, in *Ten Economic Studies in the Tradition of Irving Fisher.* Witney.

Thurow, L.C., 1969, The Optimum Lifetime Distribution of Consumption Expenditures. *The American Economic Review.* Vol. LIX. June 1969, No. 3, pp. 324–330.

Wallis, W.A. & Roberts, H.V., 1956, *Statistics. A New Approach.* Eight Printing. Glencoe.

Whelpton, P.K., 1954, *Cohort Fertility.* Princeton University Press. Princeton.

Publications in English

Statistical Methods for the Analysis of Earnings Data, with Special Application to Salaries in Swedish Industry. A. Klevmarken. (1972) 271 pp. 48:—

Mergers in Swedish Industry. An Empirical Analysis of Corporate Mergers in Swedish Industry, 1946–69. B. Rydén. (1972) 323 pp. 48:—

The Credit Market, Investment, Planning and Monetary Policy—an Econometric Study of Manufacturing Industries. G. Eliasson. (1969) 111 pp. 18:—

Publications in Swedish*

Produksjon og produktivitet i detaljhandelen (Development of the Swedish Distribution System). J. Skår. (1971) 336 pp. 25:—

Den privata konsumtionen 1931–1975 (Private Consumption in Sweden 1931–1975). C. J. Dahlman and A. Klevmarken. (1971) 158 pp. 40:—

Skoindustrins distributionsproblem (Distribution Problems of Swedish Footwear Industry). M. Lidén. (1970) 97 pp. 25:—

Svensk industri under 70-talet med utblick mot 80-talet (The Swedish Industry during the Seventies and the Outlook up to 1980). L. Nabseth. a. o. (1970) 271 pp. 40:—

Skogsindustrins strukturomvandling och expansionsmöjligheter (Forest-Based Industries: Structural Change and Growth Potentials). L. Wohlin. (1970) 291 pp. 45:—

* *Almost all studies in Swedish are published with a shoi summary in English. A list of all books published with titles in English can be obtained from the Institute on request.*

All the publications can be ordered from
Industriens Utredningsinstitut,
Box 5037, S-102 41 Stockholm 5, Sweden

In the complicated and technical process of present day collective bargaining comparisons of wages and salaries are done by more and more refined methods. Similar comparisons are central both in professional and political debate about income distributions and inequality in standard of living. Studies in the structure of earnings are also important for the analysis of consumption and savings behavior and for educational planning.

In this study two different (but related) approaches are suggested for the analysis of earnings data. The first approach is a combined time series and cross sectional analysis of age-earnings profiles by educational qualification, the second is a more detailed investigation of the earnings structure in a cross section. As an illustration they are applied to salary data from Swedish industry.

A profile for a cohort of employees is obtained as the sum of an initial salary and successive salary increments. These increments are explained by a general salary increase, an increase due to experience and an increase due to age. The applicability of the profile model is illustrated by a comparative analysis of profiles for employees with various educational qualifications, including calculations of life-time salaries.

In the cross sectional analysis salary differences are estimated due to job, branch of industry and cost of living area in addition to age and education. Furthermore, the stability of the salary structure is investigated in a comparison between four cross sections from the period 1957–1968.

In the book there are also included chapters on institutional background, labour composition and mobility and the statistical data used.

Cover design by Dick Hallström

Almqvist & Wiksell, Stockholm, Sweden ISBN 91-7204-004-1