A stencil from

IUI

On earnings-age-education profiles

by

Anders Klevmarken

THE INDUSTRIAL
INSTITUTE
FOR ECONOMIC AND
SOCIAL RESEARCH
STOCKHOLM, SWEDEN

INDUSTRIENS UTREDNINGSINSTITUT

June 30, 1971

Preliminary and Confidential

Paper to be presented at
the European Meeting of the
Econometric Society
in Barcelona 1971

ON EARNINGS-AGE-EDUCATION PROFILES

by

Anders Klevmarken

The author is research associate at The Industrial Institute for Economic and Social Research (IUI), Stockholm, Sweden

This paper is one part of a research project on the salary structure and the salary setting process in Swedish industry. Preliminary results have previously been reported in Klevmarken [1968] and [1969]. The project is supported by The National Central Bureau of Statistics, The Swedish Council for Social Science Research and The Swedish Employers' Confederation.

The author has recieved many valuable comments and suggestions to improvements on previous drafts of this paper. In particular he wishes to thank Sten Malmquist and Karl Olov Faxén, Harry Lütjohann, Hans Löfgren, Thora Nilsson, Erik Ruist, Carl-Johan Åberg and several colleagues at IUI. The author is of course responsible for remaining errors and shortcomings.

Contents

1. Introduction and summary

Earnings-age profiles are of interest in several applications. They are central in the human capital approach, 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 rate of returns to education[1)] (Blaug, Peston & Ziderman [1967]). They are also used in consumption and savings theory to explain the life cycle pattern of consumption and savings (Lydall [1955], Thurow [1969], Klevmarken [1970]). The profiles are of importance in collective bargaining. Labor unions organizing salaried employees with higher education (academic and professional educations) have in their negotiations with the employers used the characteristics of their income-age profile, low or no salary in the beginning of their active time, as an argument for high salaries compared to those groups which do not have a high education and thus have been able to earn their living relatively early. Sometimes the comparison between the groups are done on the basis of life-time incomes (SACO [1968]). The usual practice to make these calculations from cross-section profiles has been criticized (Miller [1965], Ben-Porath [1966]).

The cross-section profiles have also been used in the negotiation process in Swedish industry to determine so called statistical salary increases due to age (Lind [1963], SAF [yearly]).

During the last few years there has been an intensified debate in Sweden about individual equality with respect to income and standard of living. Comparisons between different groups have been done in several ways, but one method used is to compare income-age profiles (Eriksson [1970]). When using cohort-profiles the income increases obtained are considerably higher than those obtained from cross-section profiles, and this has created a discussion about the proper comparison (Klevmarken [1971]). The last field of application for earnings-age profiles to be mentioned is the actuarial science, where the profiles are used to estimate the money value of a man and earnings forgone because of an injury in an accident (for references see Fase [1969]).

There are two kinds of income-age profiles. One describes the individual income path through time or the income path of a cohort, the other describes the income differences between <u>different</u> individuals at

1) For a critizism see for instance Merret [1966].

one point of time. The first profile will be called cohort profile and the second cross-section profile. In most of the applications mentioned above interest is concentrated to cohort profiles and models are built to explain individual earnings, but in empirical work cross-section profiles are frequently used. Cross-section profiles are thus frequently interpreted as cohort profiles sometimes with the justification that this would be a proper procedure in a "perfectly stationary world" (see for instance Fase [1969]). As we do not live in a perfectly stationary world and as the rate of growth of the economy may effect different groups of individuals differently and thus also the shape of cross-section profiles, we may at least question the cohort interpretation of cross-section profiles and investigate under what circumstances this interpretation is justified.

The purpose of this paper is to develop a statistical framework to investigate earnings-age profiles.[1)] A model will be used which both demonstrates the differences between cohort and cross-section profiles and links them together. It is derived from a simple representation of individual earnings as the sum of initial earnings and accumulated yearly increases; initial earnings also change from one year to another. By applying different constraints to the changes of earnings various models of earnings-age profiles can be obtained. In this paper a fairly simple model is applied to between fifteen and twenty cross-sections of salary data from Swedish industry. The estimates obtained are used to calculate life-time salaries.

2. A formal representation of an individual income path

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

$$n = \sum_{t=0}^{T} n_t; \qquad (2:1)$$

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

1) A similar work has been done by Fase [1969].

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

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

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

when e_{0i} is defined such that

$$\sum_{i=1}^{n_0} e_{0i} = 0; \tag{2:3}$$

This definition implies that $\exp[\alpha]$ can be interpreted as the average[1] 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

$$(\exp[\beta_t] - 1)100; \quad t=1,\ldots,T \tag{2:4}$$

per cent from time t-1 to time t.

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

$$\exp[\alpha + \sum_{t=1}^{b} \beta_t]; \quad b=1,\ldots,T \tag{2:5}$$

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

$$\exp[\alpha + \sum_{t=1}^{b} \beta_t + e_{bi}]; \quad b=1,\ldots,T \quad i=1,\ldots,n_b \tag{2:6}$$

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

$$\sum_{i=1}^{n_b} e_{bi} = 0; \quad b=1,\ldots,T \tag{2:7}$$

---

1) Geometrical average.

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

$$(\exp[\gamma_{tb}]-1)100; \quad b=1,\ldots,T. \quad t=b+1,\ldots,T \tag{2:8}$$

per cent, and the change of the i:th individual is defined as

$$(\exp[\gamma_{tb} + u_{tbi}]-1)100; \quad b=1,\ldots,T. \quad t=b+1,\ldots,T. \quad i=1,\ldots,n_b \tag{2:9}$$

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

$$\sum_{i=1}^{n_b} u_{tbi} = 0; \quad b=1,\ldots,T. \quad t=b+1,\ldots,T \tag{2:10}$$

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

$$L_{Tb.} = \exp[\alpha + \sum_{t=1}^{b} \beta_t + \sum_{t=b+1}^{T} \gamma_{tb}]; \quad b=1,\ldots,T \tag{2:11}$$

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

$$L_{Tbi} = L_{Tb.} \exp[e_{bi} + \sum_{t=b+1}^{T} u_{tbi}]; \quad b=1,\ldots,T. \quad i=1,\ldots,n_b \tag{2:12}$$

This representation of the individual salary growth is summarized in table 2:1. 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 sequel it will be used as a starting point for model building, and this is of course the purpose of presenting this particular scheme.

Before the model work begins in section 3 it may be useful to do some preliminary exercises to examine some properties of this formal representation.

Table 2:1 Earnings paths for individuals with different periods of entry

Time

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

The average salary of all n individuals equals

$$L_{T..} = \{ \prod_{b=0}^{T} \prod_{i=1}^{n_b} L_{Tbi} \}^{\frac{1}{n}} = \{ \prod_{b=0}^{T} (L_{Tb.})^{n_b} \}^{\frac{1}{n}} =$$

$$= \exp\left[\alpha + \frac{1}{n} \sum_{t=1}^{T} (\beta_t \sum_{b=1}^{T} n_b) + \frac{1}{n} \sum_{t=1}^{T} \sum_{b=0}^{t-1} n_b \gamma_{tb}\right]; \qquad (2:13)$$

From this expression it is possible to see how the group composition and the salary changes interact and give the over all average. An average change in initial salary is weighted by the number of persons whose initial salaries depend upon this change. β_{t-1} thus obtains more weight than β_t. The changes γ_{tb} are weighted by the number of individuals who have experienced this change. To be able to say something more about these changes two alternative assumptions will be introduced. Assume first that the salary changes γ_{tb} are independent of when employment was obtained, i.e.

$$\gamma_{tb} = \gamma_t; \quad b=0,\ldots,T-1 \qquad (2:14)$$

The last term in the exponent of (2:13) then becomes

$$\frac{1}{n} \sum_{t=1}^{T} (\gamma_t \sum_{b=0}^{t-1} n_b); \qquad (2:15)$$

Again γ_t will be weighted by all those who have experienced the change γ_t, and this means that γ_t will obtain more weight than γ_{t-1}.

Now, assume alternatively that the changes γ_{tb} are independent of calendar time t, but dependent on active age (t-b). That is

$$\gamma_{tb} = \gamma_{t+k,b+k} = \gamma_{(t-b)}; \quad \text{all possible k.} \qquad (2:16)$$

The last term in the exponent of (2:13) now becomes

$$\frac{1}{n} \sum_{(t-b)=1}^{T} \gamma_{(t-b)} \sum_{b=0}^{T-(t-b)} n_b; \qquad (2:17)$$

The change experienced during the first time period on the labor market is given most weight, while the change experienced only by the oldest employees is given least weight.

Consider two individuals, who obtained employment at the same time but at two different initial salaries

$$L_{bbi} = \exp\left[\alpha + \sum_{t=1}^{b} \beta_t + e_{bi}\right]; \tag{2:18}$$

and

$$L_{bbj} = \exp\left[\alpha + \sum_{t=1}^{b} \beta_t + e_{bj}\right]; \tag{2:19}$$

respectively. Assume for simplicity that the salary changes they then obtain are equal and also that they equal the average changes of the group to which they belong. Under these assumptions the two salaries will always stand in a constant proportion to each other, namely,

$$\frac{L_{Tbi}}{L_{Tbj}} = \exp[e_{bi} - e_{bj}]; \tag{2:20}$$

The difference between the two salaries will, however, increase or decrease depending upon whether the salaries increase or decrease.

$$L_{Tbi} - L_{Tbj} = L_{Tb.} \exp[e_{bi} - e_{bj}]; \tag{2:21}$$

Assume for instance, that $e_{bi} > e_{bj}$ and that $L_{Tb.}$ increases. To compensate his low initial salary individual j needs an increase which exceeds i's increase with exactly the same percentage as i's initial salary exceeded j's. This rule holds whenever this compensation takes place. But the longer j has to wait for it, the more he must obtain expressed in crowns, dollars or any other currency.

Provided future salary increases do not depend on the initial salary obtained it is obvious that it pays very well to negotiate a high initial salary. Frequently salary increases follow a centrally negotiated agreement and the presumption is then fulfilled.

It is common to deflate salaries with a consumer's price index. The formal scheme outlined above can be used to represent salary paths expressed both in current and in constant prices.

Assume that the percentage change of the average price level, measured by some convenient index, between t-1 and t is

$$(\exp[\rho_t] - 1)100; \quad t=1,\dots,T \tag{2:22}$$

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

$$L^R_{Tb.} = L^N_{Tb.}\exp\left[-\sum_{t=1}^{T}\rho_t\right] = \exp\left[\alpha + \sum_{t=1}^{b}(\beta_t-\rho_t) + \sum_{t=b+1}^{T}(\gamma_{tb}-\rho_t)\right]; \qquad (2:23)$$

In (2:23) time period 0 is the base. Two individual salary pathes can be compared as before but now deflated to constant prices. For the example above the ratio of the deflated salaries will be the same as the ratio between the nominal salaries. The difference between the deflated salaries will increase or decrease depending on whether the deflated salaries increase or decrease.

3. Models of the salary setting process

The formal representation will now be used for model building. The model will be built up directly for application on average earnings and not on individual earnings.

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 done on a yearly (or multiyearly) basis.

To begin with a few simple assumptions will be made about the nature of average salary changes β_t and γ_{tb}.

3.1. A model with age dependent salary increases

The average initial logarithmic salary $\ln L_{bb.}$ is a stochastic variable specified as follows, where α and β are constants

$$\ln L_{bb.} = \alpha + \beta b + \varepsilon_b; \qquad (3:1)$$

$$E(\varepsilon_b) = 0; \qquad (3:2)$$

$$E(\varepsilon_b\varepsilon_{b+\tau}) = \begin{cases}\sigma^2 & \text{if } \tau = 0;\\ \rho_\tau\sigma^2 & \text{if } \tau \neq 0;\end{cases} \qquad (3:3)$$

$$[\rho_\tau] > [\rho_{\tau+1}]; \qquad (3:4)$$

The distribution of ε_b is independent of b. From (3:1) it follows that the increase of the average initial salary from year b-1 to year b, denoted by β_b, is

$$\beta_b = \beta + \varepsilon_b - \varepsilon_{b-1}; \tag{3:5a}$$

The stochastic component $\varepsilon_b - \varepsilon_{b-1}$ of this expression is below denoted by z_b, thus

$$\beta_b = \beta + z_b; \tag{3:5b}$$

The expected value of z_b is zero and its variance is constant. From (3:5) it is easy to see that

$$z_b = -z_{b-1} + \varepsilon_b - \varepsilon_{b-2}; \tag{3:6}$$

Although $\varepsilon_b - \varepsilon_{b-2}$ does not fullfil the properties of the random term in a first order Markov-chain, (3:6) indicates that a negative autocorrelation could be expected in β_b.

Already in section 2 it was indicated that the average changes γ_{tb} may vary with active age and with calendar time. The dependence on calendar time should not be understood such 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 demand and supply and to negotiations. These effects will be treated more extensively later.

It is commonly accepted that salary increases are obtained as a result of experience gained. Investment in training (on the job training) increases the marginal productivity of an employee and also earnings. 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 entrance to the labor market. During the training period marginal productivity is low and increases in earnings are thus rather low during the very first years on the labor market but then become higher as experience is gained.[1)] When investment later decreases,

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

the increase of marginal productivity and earnings diminish. The earnings profile would obtain the general shape indicated by the curve TT in figure 3:1, which is a reproduction of fig. 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.) General school education would have a similar effect on the earnings profile and in addition, 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 would then be flatter than profiles of employees with a university degree.

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

$$\gamma_{tb} = f_{\gamma'}(t-b) + v_{tb}; \tag{3:7}$$

$$E(v_{tb}) = 0; \tag{3:8}$$

Although we have an idea about the shape of the function $f_{\gamma'}$, it may be difficult to find a functional form with parameters which can easily be interpreted and also gives a good fit.[1] For this reason the age range is divided into c intervals with the limits $A_0, A_1, \ldots, A_c$ and the function $f_{\gamma'}$ is approximated by a function, which is constant within each interval. The levels of the constant segments of this function, γ_i, $i=1,\ldots,c$, can be estimated from the data. With these assumptions and with assumptions (3:2) and (3:5), (2:11) can be rewritten as a polygon

$$L_{Tb.} = \exp\left[\alpha + \beta b + \sum_{i=1}^{c} \gamma_i' D_i' + \sum_{t=1}^{b} z_t + \sum_{t=b+1}^{T} v_{tb}\right]; \tag{3:9}$$

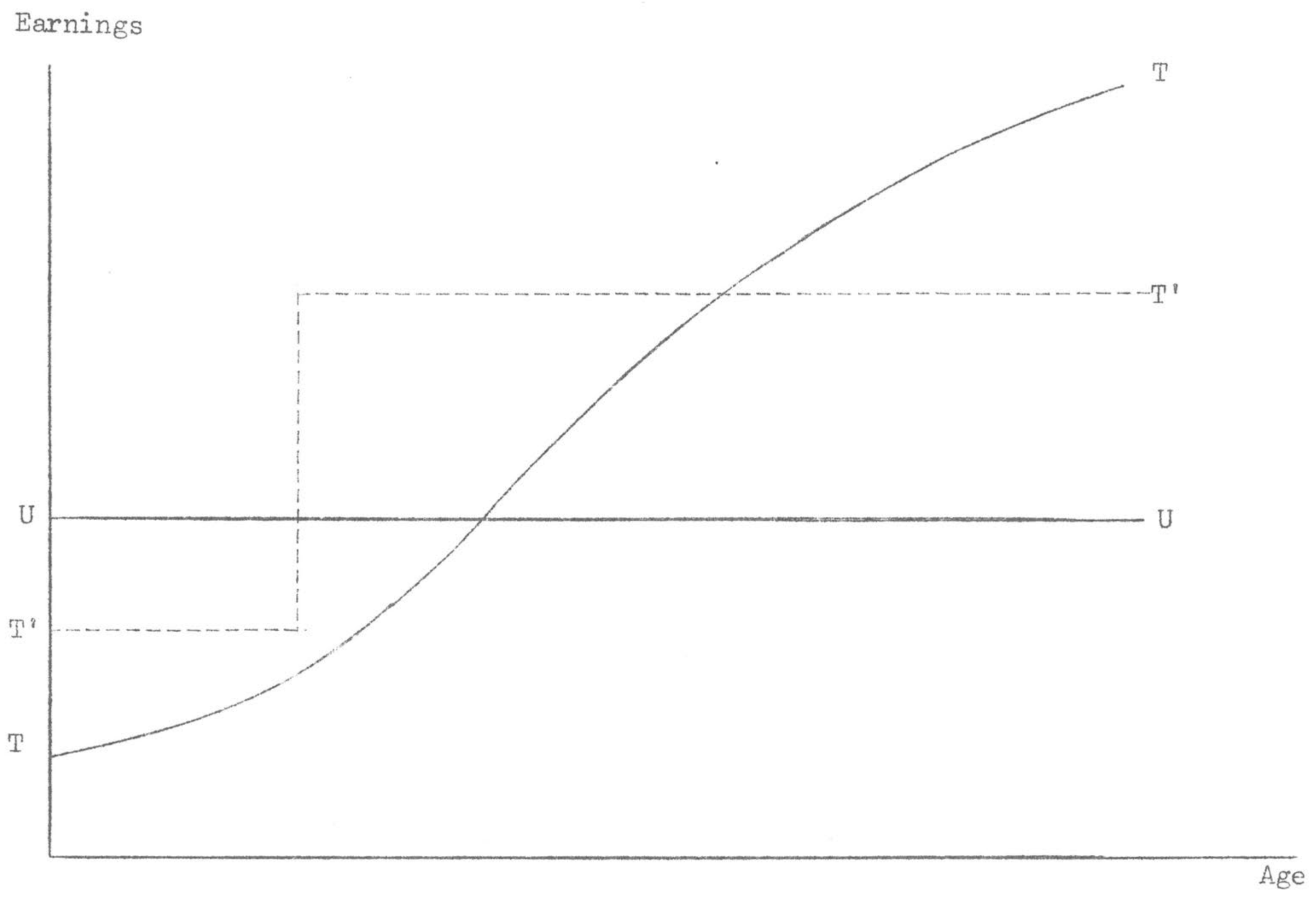
where D_i' is the time spent in active age class i or

$$D_i' = \begin{cases} 0 & T-b < A_{i-1} \\ (T-b)-A_{i-1} & \text{if } A_{i-1} \leqq T-b \leqq A_i; \\ A_i - A_{i-1} & A_i < T-b \end{cases} \tag{3:10}$$

---

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

Fig 3:1 Earnings - age profiles



Because of (3:2), (3:3), and (3:8) it is true that

$$E\left(\sum_{t=1}^{b} z_t + \sum_{t=b+1}^{T} v_{tb} \right) = 0; \tag{3:11}$$

Provided data on active age is available, (3:9) is a model which can be estimated rather easily. But before the estimation problems are approached the factors behind the increments γ_{tb} will be considered once again.

Keeping 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 is usually higher than the ability of an old employee, and this should have an influence on marginal productivity and on earnings in addition to experience. The rate of increase in earnings γ_{tb} should thus also be a function of physical age, probably a decreasing function. 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 rule in the salary setting process. As this data source gives information about average salaries in different age groups it strengthens the common practice to set salaries after 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 of notation. Those who are born year f and started to work year b will year T on the average earn L_{Tfb}, and the rate of increase between t-1 and t in their average salary is γ_{tfb}. Our new assumptions about the rate of increase in earnings can formally be written

$$\gamma_{tfb} = f_\gamma \left[(t-b), (t-f) \right] + v_{tfb}; \tag{3:12}$$

$$E(v_{tfb}) = 0; \tag{3:13}$$

$$E(v_{tfb} v_{t+i,f+j,b+k}) = \begin{cases} \sigma^2(v)_{(t-f),(t-b)} & \text{if } i=j=k=0; \\ 0 & \text{if } iUjUk \neq 0; \end{cases} \tag{3:14}$$

$$E(z_t v_{t+i,f,b}) = 0 \quad \text{all } i \tag{3:15}$$

The properties (3:14) and (3:15) of the stochastic variable v_{tfb} are not only chosen by convenience. An attempt to a justification will be made further below.

As the rate of increase in earnings do not causally depend on active and physical age, but rather on the ability to do a job, negotiation practice and other factors associated both with active and physical age, it is difficult to separate a "pure effect" of active age and a "pure effect" of physical age and specify the properties of the function f_γ. The approach taken is to postulate that the active age and the physical age effects are additive and, as before, to approximate f_γ by a polygon. The limits of the physical age intervals are the same as those of 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. The active age effect in interval i is as before denoted by γ_i' and the physical age effect is denoted by γ_i''. The new model now is

$$L_{Tfb.} = \exp\left[\alpha+\beta b+\sum_{i=1}^{c}\gamma_i'D_i' + \sum_{i=1}^{c}\gamma_i''D_i'' + \sum_{t=1}^{b} z_t + \sum_{t=b+1}^{T} v_{tfb}\right]; \qquad (3:16)$$

D_i' is already defined by (3:10) and D_i'' is the time spent in physical age class i or

$$D_i'' = \begin{cases} 0 & y < A_{i-1} + C; \\ y-(A_{i-1}+C) & \text{if } C + A_{i-1} \leq y \leq A_i + C; \\ A_i - A_{i-1} & y > A_i + C; \end{cases} \qquad (3:17)$$

From (3:2), (3:5) and (3:13) it follows that

$$E(\sum_{t=1}^{b} z_t + \sum_{t=b+1}^{T} v_{tfb}) = 0; \qquad (3:18)$$

To use active age as an independent variable is troublesome from empirical point of view. Salary statistics is usually not collected with 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 passed their exams. Most civil 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 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? When a person starts to work this is of course first of all determined by his birth date, but education, military service, illness and other personal matters also influence time for the 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 conveniently narrowly defined group of employees. As education seems to be the factor differentiating most, it is natural to consider individuals with the same education.

If the age when employment is obtained for the first time is denoted by y_b, the year of birth by f and the physical age by y as before, the following two identities hold true for all individuals

$$T-b \equiv y - y_b; \tag{3:19}$$

$$b \equiv f + y_b; \tag{3:20}$$

y_b will be considered as a stochastic variable, while f is not. The properties of the distribution of y_b given f and the distribution of y_b given b will be discussed later.

Suppose the model (3:16) is applied to such a homogeneous group of employees that every member of the group started to work at the same age, $y_b = C$. By substitution of (3:19) and (3:20) into (3:16) this expression can now be simplified to

$$L_{Tf.} = \exp\left[\alpha'+\beta f + \sum_{i=1}^{c} \gamma_i D_i + \sum_{t=1}^{b} z_t + \sum_{t=b+1}^{T} v_{tf}\right]; \tag{3:21}$$

where

$$\alpha' = \alpha + \beta C; \tag{3:22}$$

$$\gamma_i = \gamma_i' + \gamma_i'' ; \tag{3:23}$$

$$D_i = D_i' = D_i'' \tag{3:24}$$

Average earnings are now a function of date of birth and physical age. There would thus be no problem if all individuals started to work at the same age, but this is not so even for a relatively homogeneous educational group like civil engineers. If this strong assumption is not fulfilled, how serious is then the specification error committed when (3:21) is used?

In order to give some insight into this problem a simplified example will be investigated. Assume the "normal" age when employment is first obtained is C years and that deviations in either direction may occur with one year only. Suppose in addition that the age intervals used in model (3:13) have a length of one year. Among those who in year b are C years old there will be one group of say n_1 employees who started to work already in year b-1 and one group of say n_2 employees who started to work in year b. The average salaries these two groups obtain are given in column two of tables 3:1A and B. The two groups can be followed one year and their salaries year b+1 are found in the third column of the same two tables. During this year, b+1, a third group of n_3 employees join labor market and their average salary is also given in the third column. The mean salary by physical age is given in the bottom row of the tables. A comparison with (3:21) shows that this model contains a systematic error. From the age C+1 on the importance of the error depends on the relative magnitudes of γ_i' and γ_{i-1}' and on the distribution of y_b given f. Suppose n_1 and n_3 are at least approximately equal, and each about 25 % of the whole group. The error then mainly depends on the difference $\gamma_i' - \gamma_{i-1}'$ which will probably be at most a few percent of the average salary. The total error would then stay well below one percent. Remembering figure 3:1 we would expect the error to be positive for young employees and negative for old employees. If (3:21) is applied the <u>estimates</u> of γ_i would then tend to get a negative bias. These conclusions are based on the assumption of a symmetric and stable group composition. The distribution of y_b given f is usually positively skew because it is easier to take another year in school than to speed up one year. The effect of this skewness would be that the positive systematic error among young employees is reduced and the negative among old employees is strengthened. In practical work it is usually not possible to follow exactly the same individuals over a life-time. Some individuals will leave the group and others will join. The constance of the relative magnitudes of the number n_i will then hold only approximately.

Because y_b is a stochastic variable there is another problem to consider. The numbers n_i are stochastic because y_b is stochastic with different realizations for each f. As f is used in the model (3:21) as an independent variable it is a necessary requirement of the usual linear

Table 3:1A <u>Parameter representation of logarithmic earnings of a cohort by active and physical age</u>

Active age	Physical age (and time)		
	$c-1(b-1)$	$c(b)$	$c+1(b+1)$
0	$\alpha+\beta(b-1)$	$\alpha+\beta b \quad +\gamma''_0$	$\alpha+\beta(b+1)+\gamma''_0+\gamma''_1$
1		$\alpha+\beta(b-1)+\gamma''_0+\gamma'_1$	$\alpha+\beta b \quad +\gamma''_0+\gamma''_1+\gamma'_1$
2			$\alpha+\beta(b-1)+\gamma''_0+\gamma''_1+\gamma'_1+\gamma'_2$
Mean	$\alpha+\beta(b-1)$	$\alpha+\beta b+\gamma''_0+\dfrac{n_1}{n_1+n_2}(\gamma'_1-\beta)$	$\alpha+\beta b+\gamma''_0+(\gamma''_1+\gamma'_1)+\dfrac{n_1(\gamma'_2-\beta)-n_3(\gamma'_1-\beta)}{n_1+n_2+n_3}$
	$\alpha'+\beta(f-1)$	$\alpha'+\beta f+\gamma''_0+\dfrac{n_1}{n_1+n_2}(\gamma'_1-\beta)$	$\alpha'+\beta f+\gamma''_0+(\gamma''_1+\gamma'_1)+\dfrac{n_1(\gamma'_2-\beta)-n_3(\gamma'_1-\beta)}{n_1+n_2+n_3}$

Table 3:1 B Residual specification of logarithmic earnings of a cohort by active and physical age

Active age	Physical age (and time)			
	C-1(b-1)	C(b) C(b)	C+1(b+1)	C+2(b+2)
0	$\sum_{t=1}^{b-1} z_t$	$\sum_{t=1}^{b} z_t + v_{bb}$	$\sum_{t=1}^{b+1} z_t + v_{(b+1)(b+1)} = R_{(b+1)(b+1)}$	
1		$\sum_{t=1}^{b-1} z_t + v_{b(b-1)}$	$\sum_{t=1}^{b} z_t + v_{bb} + v_{(b+1)b} = R_{(b+1)b}$	$\sum_{t=1}^{b+1} z_t + v_{(b+1)(b+1)} + v_{(b+2)(b+1)} =$ $= R_{(b+1)(b+1)} + v_{(b+2)(b+1)}$
2			$\sum_{t=1}^{b-1} z_t + v_{b(b-1)} + v_{(b+1)(b-1)} = R_{(b+1)(b-1)}$	$\sum_{t=1}^{b} z_t + v_{bb} + v_{(b+1)b} + v_{(b+2)b} =$ $= R_{(b+1)b} + v_{(b+2)b}$
3				$\sum_{t=1}^{b-1} z_t + v_{b(b-1)} + v_{(b+1)(b-1)} + v_{(b+2)(b-1)} =$ $= R_{(b+1)(b-1)} + v_{(b+2)(b-1)}$
Mean	$\sum_{t=1}^{b-1} z_t$	$\sum_{t=1}^{b-1} z_t + \frac{n_2(z_b + v_{bb}) + n_1 v_{b(b-1)}}{n_1 + n_2}$	$\frac{n_1 R_{(b+1)(b-1)} + n_2 R_{(b+1)b} + n_3 R_{(b+1)(b+1)}}{n_1 + n_2 + n_3}$	$\frac{n_1 R_{(b+1)(b-1)} + n_2 R_{(b+1)b} + n_3 R_{(b+1)(b+1)}}{n_1 + n_2 + n_3}$ $+ \frac{n_1 v_{(b+2)(b-1)} + n_2 v_{(b+2)b} + n_3 v_{(b+2)(b+1)}}{n_1 + n_2 + n_3}$

The short form R_{Tb} is introduced only to simplify the notation. As the table gives the earnings paths of one cohort only the sub-index f is dropped.

Table 3:1C Parametric representation of logarithmic earnings of a labor market cohort[1] by active and physical age

Active age (and time)	Physical age C-1	C	C+1	C+2	Mean
0 (b)	$\alpha+\beta b$	$\alpha+\beta b+\gamma''_0$	$\alpha+\beta b+\gamma''_0+\gamma''_1$		$\alpha+\beta b+\gamma''_0 + \frac{m_3\gamma''_1-m_1\gamma''_0}{m_1+m_2+m_3}$
1 (b+1)		$\alpha+\beta b+\gamma''_0+\gamma'_1$	$\alpha+\beta b+\gamma''_0+\gamma''_1+\gamma'_1$	$\alpha+\beta b+\gamma''_0+\gamma''_1+\gamma''_2+\gamma'_1$	$\alpha+\beta b+\gamma''_0+(\gamma'_1+\gamma''_1)+\frac{m_3\gamma''_2-m_1\gamma''_1}{m_1+m_2+m_3}$

1) Labor market cohort = those who enter labormarket at the same time.

model that the distribution of y_b is independent of f.[1)] The average age of a person who enters labor market may not be constant, for instance, due to changes in the educational system, to changes in the length of military service, to changing labor market conditions and so forth. As mentioned before there are no data on active age and it is not possible to investigate directly the age distribution of those who obtain their first job. However, there are some data on examination age, which makes it possible to study the age distribution of those who obtain a degree or a certificate. The mean and standard error of some empirical distributions are given in table 3:2. It should be observed that the table gives statistics of distributions by calendar year (b) rather than by cohorts (f). These numbers should be compared with some care, because they are obtained by a number of approximations and the comparability is also restricted due to shifting definitions. With exception of 1969/70 the age data for the university degrees refer to the time point when the degree was obtained. The age of those who got their degree 1969/70 is measured 1970, which yields an overestimate of the age at the examination. The high school figures are obtained during the fall of the last year in school, while degrees usually are granted at the end of the spring semester.

The mean examination age has been very stable for the university degrees during the sixties. Also the high school degrees exhibit a stable average examination age. A small decline may be noticed for engineering I and an increase for engineering II. The age dispersion is a little higher among science graduates, and technicians (engineering II). For this last group the standard deviation is probably underestimated because no data are available for part time students who are frequent in engineering II and usually older students. The high standard deviation among science graduates should not worry too much. Most of these students do not go to industry and it seems to be a fair guess that those who do go are a more homogeneous group.

In conclusion the results from table 3:2 indicate that the assumption of independence between f and the distribution of y_b is rather harmless.

---

1) The present situation is not an analogy to the usual "errors in variable" case but rather to the Berkson case (Berkson [1950]).

Table 3:2 Average age when a degree is obtained

Education		1961/62	1962/63	1963/64	1964/65	1965/66	1966/67	1969/70
University degrees								
Engineering	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	X			26.8				26.3
(Civilekonom)	S			2.4				2.9
Science (Naturvetare,	X			26.2				26.1
FK, FM)	S			3.8				3.5
High School degrees								
Engineering I (Läroverksin-	X						20.7	20.1
genjör)	X						2.4	1.7
Engineering II (Institutsin-	X						23.0	24.2
genjör)	S						3.2	3.9
Business (Gymnasie-	X			20.8			20.1	19.5
ekonom)	S			2.5			1.8	2.5

Data sources and comments

University degrees

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

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

High school degrees

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

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

The model (3:9) can also be considered as a simplified version of (3:16). The choice between the three models (3:9), (3:16) and (3:21) is mainly determined by the data available. A comparison between the models (3:9) and (3:16) can be done in exactly the same way as the comparison between (3:21) and (3:16). Table 3:1C contains a parametric representation of logarithmic earnings of employees who obtained their first job at the same time (a labor market cohort). The same example is used as before, i.e. three cohorts make one labor market cohort. Averaging over cohorts gives the earnings-active age profile in the last column of table 3:1C. A comparison with expression (3:9) reveals a systematic specification error of the same kind as before. The greater differences between the physical age effects $\gamma''_i - \gamma''_{i-1}$ the greater error. According to our a priori notion about the physical age effect γ''_i should be smaller than γ''_{i-1}, but this difference is perhaps balanced by the positive skewness of the distribution of y_b. The specification error in (3:9) should thus be small.

Before any of the models (3:9), (3:16) or (3:21) can be estimated their stochastic properties have to be examined in more detail. The stochastic variables z and v are deviations of the <u>average</u> rate of increase in earnings from the expected increase. The residual variance depends both on the individual variability in earnings and on the number of individuals forming an average. We thus do not only look upon z and v as "unexplained residuals" although imperfections of the model certainly contribute to the residual variance too.[1)] In order to give the model plausible stochastic properties the individual variability of earnings will first be discussed and then the averaging procedure.

Empirically we know that the higher average salary and the older employees the greater dispersion (Hill [1959], Morgan [1962]). One important explanation to this is that most employees start on about the same job level at approximately the same salary. Some then obtain promotion to more responsible and higher payed jobs, while others do not, or at least not as quickly. This suggests that the variance of salary increases is highest in the age class where promotion differentiate the salary increases most. During the first five to ten years on the labor market most employees 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

1) For the moment we disregard from short-run effects on the rate of increase in earnings which are not explained by the model and which work more or less on each individual increase.

available on top level. Therefore, a plausible hypothesis is that the variance for this reason is highest in the age class 35-45 years.

The published salary statistics from the Swedish Employers' Corporation admit a rough check of this hypothesis. Table 3:3 contains semi-interquartile ranges calculated from the published tables of civil engineers and high school engineers. They exhibit a variability which clearly increase by age. However, a warning must be given against this interpretation. The age intervals used differ in length and as the average salary increases by age, the dispersion would be higher in the wide intervals than in the narrow intervals. As wide intervals are used at the end of the age distribution table 3:3 may exaggerate the increase in dispersion. Furthermore as the average rate of increase in salary differs from age interval to age interval, measures of variability from intervals of equal length would not be perfectly comparable, and even if the average rate of increase was constant over the whole age range the semi-interquartile ranges calculated from grouped data would overestimate the individual dispersion. The measures of dispersion in table 3:3 need to be standardized for these effects before a proper comparison can be made.

Denote the standard deviation of individual logarithmic salaries of those who are y years old σ_y, the standard deviation in age interval i S_i, and the expected value $E(L_{Tf.}) = \mu_{Tf}$. Assume that the age distribution <u>inside</u> each age interval is uniform. The variance of logarithmic salaries in an age interval n years wide, $S^2_{i(n)}$, can then be written as follows

$$S^2_{i(n)} = \frac{1}{n} \sum_y \sigma^2_y + \sum_{k\ell}\sum \left(\frac{\mu_{Tk} - \mu_{T\ell}}{n}\right)^2; \qquad (3:25)$$

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 $(\mu_{Tk} - \mu_{T\ell})\frac{1}{n}$. From (3:21) it follows that

$$\mu_{Tk} - \mu_{T\ell} = (\gamma_i - \beta)(k - \ell); \qquad (3:26)$$

The estimates of γ_i and β obtained below can now be used to estimate the second term of (3:25). The results from these calculations are presented in table 3:4. The variance component due to unequal age intervals and salary increases is small compared to the total variance. The remaining individual variability increases by age. In figure 3:2 the variances due to individual variablity are plotted against age. There is a small indication of a smaller increase in variability between 30 and 40 years for civil engineers and a few years earlier for highschool engineers, and a higher increase than average immediately before and after this age interval. This indicates then a smaller variance of salary increases, $\sigma^2(v)$, between 30 and 40 years (25 and 35 for highschool educated employees) and a higher variance among younger and older employees.

Table 3:3 Semi-interquartile ranges of logarithmic salaries

Education/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-
Civil engineers												
1956			.039	.071	.106	.113	.120	.138	.166	.210	.216	.292
1962			.039	.071	.092	.104	.115	.117	.140	.150	.180	.237
1969			.051	.062	.104	.104	.113	.117	.127	.134	.164	.196
High school engineers I												
1956	.037	.067	.067	.104	.111	.111	.127	.150	.182	.184	.203	.235
1962	.044	.060	.067	.083	.092	.104	.117	.129	.152	.173	.187	.242
1969	.058	.067	.092	.101	.104	.115	.117	.140	.140	.152	.182	.212

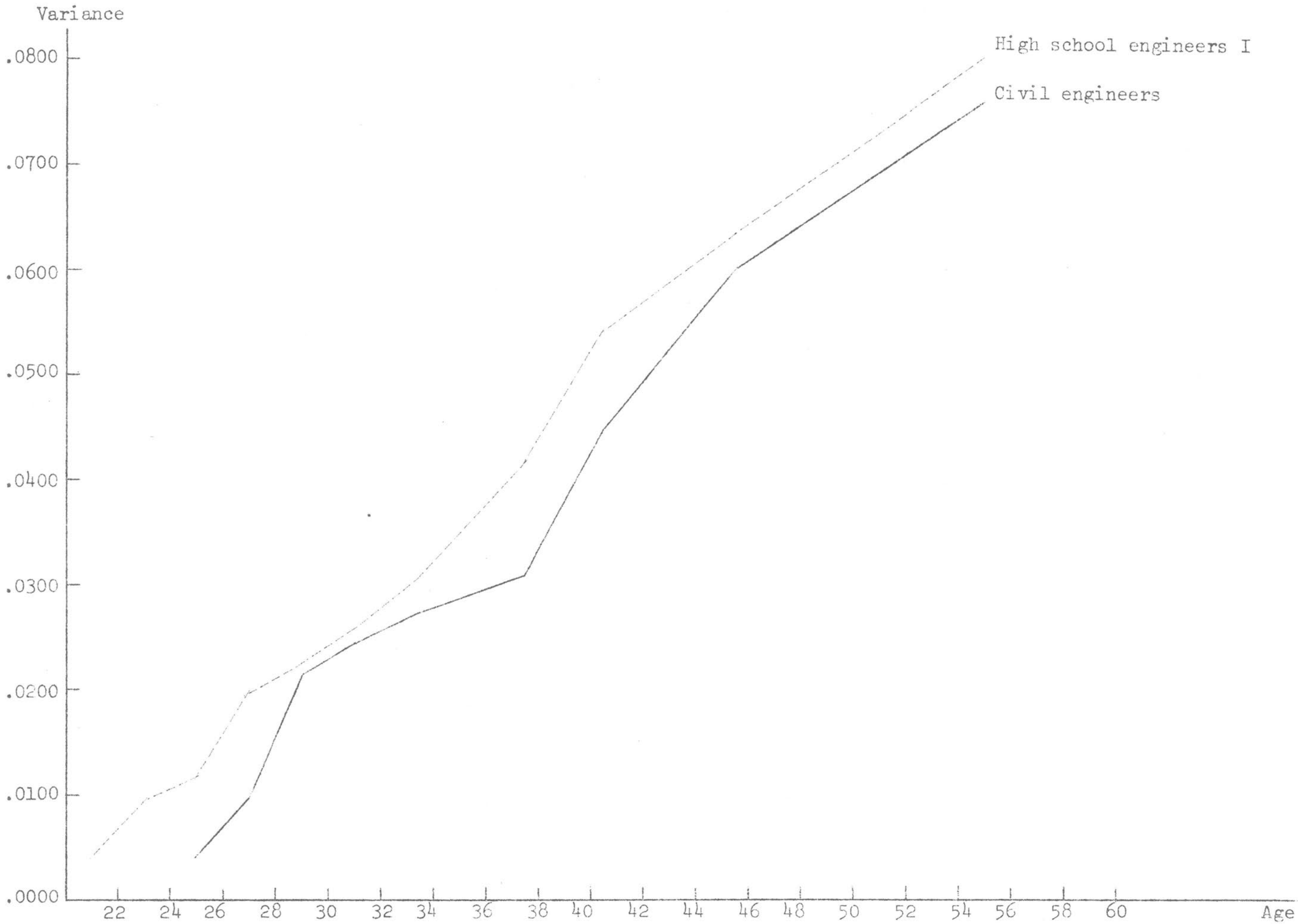
Note: The semi-interquartile range is obtained as 1/2 $(\ln Q_3 - \ln Q_1)$, where Q_1 and Q_3 are the first and the third quartile salary.

Table 3:4 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
Civil engineers											
Variance due to unequal age intervals and salary increases			.0002	.0006	.0008	.0009	.0020	.0029	.0008	.0002	.0007
Variance due to individual variability			.0039	.0096	.0217	.0244	.0276	.0310	.0446	.0598	.0760
Standard deviation ditto			.062	.098	.147	.156	.166	.176	.211	.245	.276
High school engineers I											
Variance due to equal age intervals and salary increases	.0002	.0003	.0005	.0006	.0006	.0005	.0009	.0015	.0006	.0000	.0002
Variance due to individual variability	.0044	.0089	.0118	.0196	.0222	.0261	.0308	.0418	.0542	.0635	.0799
Standard deviation ditto	.066	.094	.109	.140	.149	.162	.175	.204	.233	.252	.283

Note: The results in this table are obtained by an approximation. S_i in expression (3:25) is an average of three semi-interquartile ranges from table 3:3, divided by the factor .6745, which is obtained from a table of the normal distribution.

Fig. 3:2 Individual salary variability by physical age



Whether this dependence on age should properly be specified as a dependence on active or physical age is a difficult question. To keep all possibilities open the variance in (3:14) is indexed by both variables. The second half of assumption (3:14) implies that successive disturbances are uncorrelated. The existance of compensatory demands is perhaps an argument in favour of a negative correlation but there are no strong evidence supporting this hypothesis and it would complicate the model.

The more employees there are in an age interval the smaller will the residual variance be. To give some indication of the age (physical age) distribution inherent in the data used, the distributions from 1962 are presented by education in table 3:5. They do not differ very much from education to education. The model age interval is 35-44 years for all educations. It may also be mentioned that the total number of employees increases by time (see table 3:6). There would thus be a tendency to decreased variability by time.

The implications for average earnings of the stochastic specification now discussed at some length can be read from (3:16) and table 3:1B. The variance of logarithmic earnings will increase by age. Because of the increasing number of employees there may be a tendency to decreased variability by time. From (3:16) it can also easily be seen that for any combination of f and b the residuals of the model follow a first order Markov chain with the proportionality coefficient equal to one. In table 3:1B it is shown that the same is true for the residuals of (3:21) given f, with exception of the very first age intervals. (Nor does it hold for the last intervals.)

In section 2 $L_{Tb.}$ was defined as a geometrical average. However, in our data sources only median salaries are tabulated. 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 of a cohort can be approximated by this distribution. If a 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 wellknown properties of the lognormal distribution it then follows that $\exp[\mu]$ is the median of the lognormal distribution, while the expected value is $\exp\left[\mu + \frac{\sigma^2}{2}\right]$.

Table 3:5 Age distributions 1962

Age classes	Number of employees in SAF						
	-21	22-25	26-29	30-34	35-44	45-59	60-
University education							
Engineering	-	112	697	1 146	1 688	909	233
Business	-	19	161	248	368	161	19
Science	-	6	20	70	99	55	11
High school education							
Engineering I	137	1 363	1 748	2 188	3 800	2 177	446
Engineering II	100	1 217	2 189	2 939	7 440	3 764	354
Business	52	396	535	628	914	640	116

Table 3:6 Number of young employees 1954-1956 and 1969

Age class	-21			-25		
	1954	1956	1969	1954	1956	1969
University education						
Engineering				33	78	155
Business				13	11	79
Science				-	6	12
High school education						
Engineering I	14	60	205			
Engineering II	20	63	76			
Business	-	31	167			

If L_{Tbi} follows a lognormal distribution with parameters μ and σ, then the average $L_{Tbi.}$ also follows a lognormal distribution with expected value $\exp\left[\mu + \frac{\sigma^2}{2n_b}\right]$. 1) For large n_b the difference between the median salary and this expected value is small and the error introduced when using median salaries instead of geometrical averages is probably small too.

Alternatively, to use median salaries can be looked upon as a redefinition of $L_{Tb.}$. If $\left[e_{bi} + \sum_{t=b+1}^{T} u_{tbi}\right]$ in (2:12) is considered as a stochastic variable which follows a normal distribution with expectation zero, L_{Tbi} follows a lognormal distribution with the median $L_{Tb.}$. There is thus some justification for using median salaries in estimating the model.

The final bias source to be mentioned is a selection in the data collection. The statistics from SAF do not cover employees on the top management level. Those older employees who remain in the statistics are thus those who have not got a promotion. The practice not to submit salaries of employees on management level may in small companies even result in a drop out on 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 is no statistics collected to exhibit the salaries on the management level and it is therefore difficult to know how important this selection bias is. It should be most severe among civil engineers and business economists, which are the most frequent group on management level, and less severe among high school educated employees. An attempt has been made in Klevmarken [1968] to use statistics from the labor union organizing civil engineers 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 of earnings are also a few percentages higher in the age bracket 35 years and on than the estimates on SAF-data given below.

---

1) y is independently $N(\mu,\sigma)$

$X = e^y$ is then $\log N(\mu,\sigma)$, and $E(X) = \exp\left[\mu + \frac{\sigma^2}{2}\right]$;

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

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

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

The model (3:16) is estimated by the ordinary least-squares method without paying any regards to the particular stochastic features of the model. The estimates from SAF-data are presented in table 3:7. The estimates of the average increase in initial salaries reveal a remarkable similarity between the educations. For all educations this increase is estimated to 6-7 % per year. 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. The salary increases of the oldest are even lower than for the youngest, but this may be the result of the selection pointed out before.

A comparison between the educations reveals that the profiles of the three university educations engineering, business and science are very similar. Science exhibits a minor deviation in the age class 30-34 years. The smaller average increase for science graduated 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 educations exhibit some dissimilarities. The increases obtained by those who have business 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.

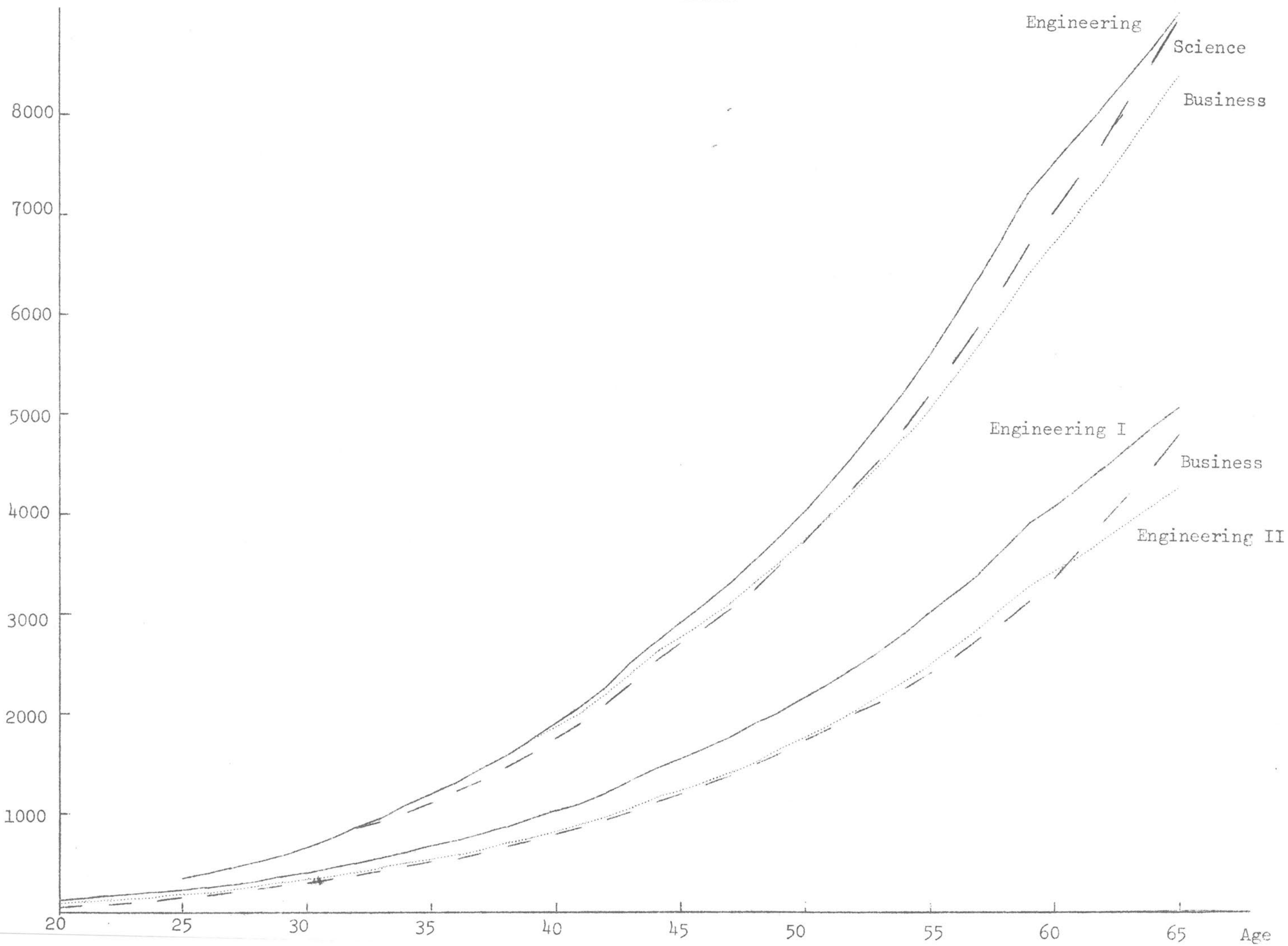
In figure 3:3 estimated cohort profiles of employees born 1910 are drawn. It is assumed that those who are graduates from a university started to work at the age of 25 and those who are graduates from highschool started to work at the age of 20 years. All the profiles illustrate salary paths in current prices and they show no decrease in any age class, on the contrary, as we have found the profiles are rather of the increasing exponential type. The diagram also clearly shows the increasing inequality by age between those who have a university degree and those who have not.

The formal representation in section 2 shows that increases in earnings could in principle vary by calendar time. In this section the dependence on active and physical age has been stressed and the dependence on calender time has been neglected. None of the models suggested catch effects of demand and supply and of negotiations, and these factors may have effects more associated with calendar time than with age. No attempt will now be made to explain this in detail, but a few remarks may be of some interest.

Table 3:7. Least-squares estimates of model (3:21) on SAF-data

Education			Age interval								
			-21	22-25	26-29	30-34	35-44	45-59	60-		Sample period
	α'	β	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	R	
University education											
Engineering	4.178	.067	-	.087	.124	.127	.091	.065	.037	.9971	1954-1969
		(.001)		(.005)	(.004)	(.003)	(.002)	(.001)	(.005)		
Business	4.240	.066	-	.070	.134	.125	.086	.060	.045	.9910	1952-1969
		(.001)		(.013)	(.006)	(.005)	(.002)	(.002)	(.009)		
Science	4.263	.062	-	.106	.125	.099	.091	.065	.048	.9779	1956-1969
		(.002)		(.040)	(.011)	(.007)	(.004)	(.004)	(.014)		
High school education											
Engineering I	3.620	.065	.085	.107	.116	.104	.086	.066	.044	.9958	1952-1969
		(.001)	(.006)	(.004)	(.004)	(.003)	(.002)	(.002)	(.006)		
Engineering II	3.339	.067	.109	.111	.115	.100	.085	.070	.043	.9948	1952-1969
		(.001)	(.006)	(.005)	(.004)	(.004)	(.002)	(.002)	(.007)		
Business	3.224	.065	.129	.138	.124	.101	.085	.066	.071	.9944	1956-1969
		(.001)	(.007)	(.006)	(.005)	(.004)	(.002)	(.002)	(.008)		

Fig. 3:3 Estimated cohort profiles by education



The supply of educated labor in Sweden has increased during the whole sample period but most so during the last few years. While there are no indications on short-run fluctuations of supply, demand show pronounced fluctuations. The effect of both demand and supply on earnings should be most observable for young employees. When demand is high relative to supply earnings of young employees would increase more than earnings of middle aged and old. The earnings-age profiles would then become flatter. The opposite should be true in the case of excess supply. Table 3:8 shows the number of jobs available for engineers and technicians advertised in Dagens Nyheter, which is a measure of demand for this kind of labor. There are peaks 1954, 1960, 1965 and 1969 and there are troughs 1958, 1962 and 1967 all in close agreement with the general business cycle in Sweden. A comparison with the residuals from the estimated model (3:21) for civil engineers and highschool engineers I in table 3:9 does not reveal any systematic positive association between demand for labor and earnings, not even so for young employees. This preliminary study thus do not 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 be investigated at the same time as those already mentioned.

3.2 Cohort-profiles and Cross-section Profiles

For each f the model (3:16) describes a cohort profile as a function of c and D_i, but by a simple transformation it can also be used to describe a cross-section profile. To simplify the notation the stochastic residuals are disregarded from. If the logarithmic residuals of (3:16) are normally distributed with expectation zero and constant variance, the non-stochastic part of (3:16) is a median salary (see above).

Table 3:8. Selected labor market indicators

Year	Number of jobs for engineers and technicians advt. in Dagens Nyheter	Percentage increase in productivity	Percentage increase in consumers' price index	Percentage increase in negotiated salary
1954	14,587	6.4	.8	.0
1955	14,444	1.9	3.1	9.1
1956	10,028	5.5	4.5	4.0
1957	10,713	4.8	4.3	3.7
1958	9,647	4.7	4.8	3.5
1959	13,047	6.8	.7	1.9
1960	15,596	5.0	3.9	3.8
1961	14,567	4.1	2.5	3.5
1962	12,953	5.4	4.3	4.7
1963	14,504	7.0	2.9	3.5
1964	15,888	8.0	3.4	4.0
1965	17,676	7.5	5.0	3.3
1966	12,009	5.0	6.3	7.2
1967	8,319	8.1	4.5	5.4
1968	9,252	8.8	1.9	5.0
1969	14,593	9.0	2.8	3.5

Table 3:9 Logarithmic residuals from model (3:21) by education, physical age and time

Year	Civil engineers			High school engineers I			University degree in business			High school certificate in business		
	25	42.5	55	21	37.5	55	25	42.5	55	21	37.5	55
1954	-.006	.033	-.056	-.067	-.030	-.037	-.051	.062	.019	-	-	-
1955	-.031	.053	.005	-.033	.015	.005	-.054	.027	-.004	-	-	-
1956	-.018	.040	-.010	.020	-.004	-.003	-.018	.058	-.028	.014	-.021	.074
1957	.021	.025	-.029	.061	-.010	-.003	-.048	.078	-.098	-.037	-.000	-.004
1958	-.009	.013	-.012	.045	-.028	-.021	-.044	-.021	-.070	-.004	-.006	-.030
1959	-.036	-.007	-.052	.027	-.050	-.047	-.001	-.046	-.103	-.022	-.026	-.024
1960	-.018	-.022	-.039	.050	-.047	-.039	-.086	-.034	-.085	-.032	-.034	-.022
1961	.039	.019	.039	.084	.012	.010	.008	.016	-.008	.022	.014	.007
1962	.059	.017	.053	.074	.029	.031	.061	.013	.029	.029	.017	.044
1963	.061	.016	.038	.079	.022	.030	.038	-.013	.060	.022	.004	.021
1964	.050	.004	.029	.047	.017	.010	.023	-.028	.087	.056	-.011	.017
1965	.062	-.001	.001	.021	.025	.006	.044	-.020	.061	.023	.010	-.004
1966	.044	-.007	.031	.010	.037	.026	.030	-.001	.048	.071	.027	.007
1967	.028	-.013	.024	-.000	.023	.012	.024	.010	.025	.035	.038	.008
1968	-.021	-.041	.010	-.090	.009	.010	.014	.008	.018	-.030	.026	.002
1969	-.076	-.069	-.016	-.121	-.024	-.005	-.034	-.016	-.022	-.031	-.036	-.081

$$M(L_{Tf.}) = \exp\left[\alpha' + \beta f + \sum_{i=1}^{c} \gamma_i D_i\right]; \tag{3:27}$$

The symbol M is used as an operator for the median. Substitution of the identity

$$f \equiv T-y; \quad y \geq C \tag{3:28}$$

and

$$y = C + \sum_i D_i \tag{3:29}$$

yields

$$M(L_{Tf.}) = \exp\left[\alpha + \beta T + \sum_{i=1}^{c} (\gamma_i - \beta) D_i\right]; \tag{3:30}$$

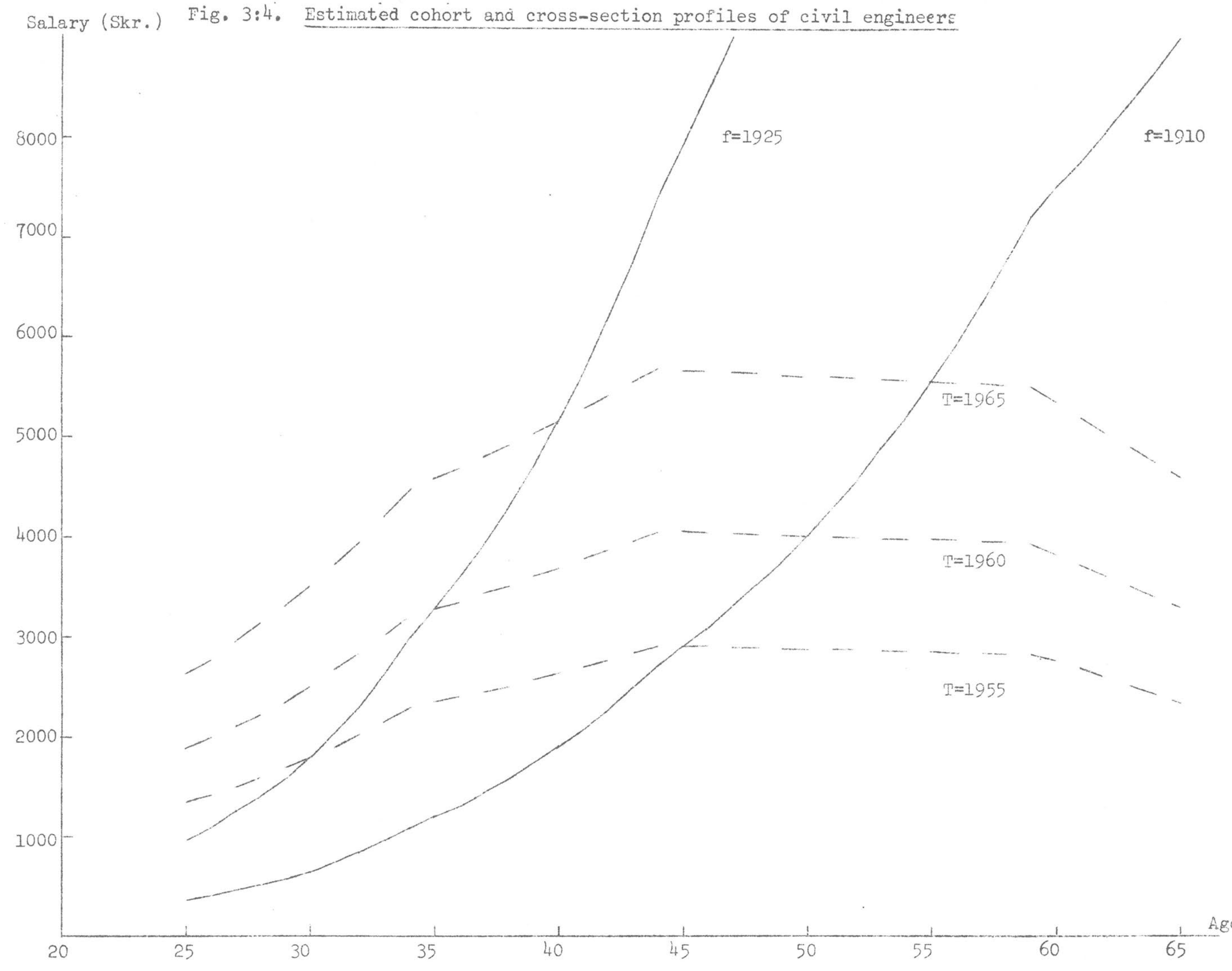
Given T, (3:30) describes a cross-section profile as a function of c and D_i. The last term in the exponent determines the shape of the profile. As long as the yearly increases γ_i are greater than the increases in initial salary, β, the profile increases, but when the opposite is true, the profile turns down. As the results in table 3:7 show old employees do not obtain increases as high as the increases of initial salaries which is an explanation of the downturning cross-section profile. It is important to notice that this downturn does not necessarily imply that any employee has got any salary decrease. This is illustrated in figure 3:4 with data for civil engineers. (3:30) also gives a new interpretation to β, namely the average yearly shift of a cross-section profile.

Suppose two groups of employees I and II, for instance two educational groups, have the same salary increases due to age, $\{(\gamma_i-\beta)\}$, but different initial salaries. The salary structure of the two groups can then be written.

$$M(L_{Tf.}) = \exp\left[(\alpha_I + \beta_I T)X_I + (\alpha_{II} + \beta_{II} T)X_{II} + \sum_{i=1}^{c} (\gamma_i - \beta) D_i\right]; \tag{3:31}$$

where X_I and X_{II} are two dummy variables which take the values 1 and 0 depending on whether an observation belongs to group I or II. The ratio between the two expected salaries thus decrease or increase depending on the increase of the initial salaries. As the results in table 3:7 indicate there is no difference in initial salary increase between educations (but some small differences in increases by age), the model may be reformulated once again.

Fig. 3:4. Estimated cohort and cross-section profiles of civil engineers



$$M(L_{Tf.}) = \exp\left[\omega_0 + \sum_{i=I}^{II} \omega_i X_i + \sum_{i=1}^{c} (\gamma_i - \beta) D_i\right]; \qquad (3:32)$$

where ω_0 is the parameter of an over all average salary a given year and ω_i, i=I,II, are deviations from the average due to group (education). Models of this type have been estimated by for instance Hill [1958], Klevmarken [1968] and Holm [1970].

3.3 Life-time salaries

The life-time salary of one employee is the sum of all his salaries earned during his active time on the labor market. The expected life-time salary of a group of individuals with a common characteristic, for instance the same education, can be estimated by a sample mean of life-time 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 use the current salaries on each job level (see for instance SACO [1968]). Frequently life-time salaries (earnings) are calculated as the area under 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 to use the estimates of model (3:16) and to calculate sums of estimated median salaries, (to be distinguished from the median life-time salary). In table 3:10 estimates of this type are given for employees who are born 1969. The calculations are made under the assumption 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 highschool certificate.

Three alternative discount rates are used, 5 %, 10 % and 15 %. All life-time salaries are discounted to the age of 20 years. The choice of discount rate greatly effects the life-time salaries in Swedish crowns, but also the relative differences between highschool educations on one side and university educations on the other, as shown by the index numbers. Inside each of the two groups the relative salaries are not very sensitive to changes of the discount rate. The relative salary of civil engineers decreases a little as the discount rate increases, and the relative salary of employees with a highschool certificate in business increases a little.

Table 3:10 Estimated lifetime salaries by education

Education	Lifetime salaries (thous.Skr)						Active time
	Discount rate 5 %	Index	Discount rate 10 %	Index	Discount rate 15 %	Index	
University education							
Engineering	19.086	240	5,313	219	1.977	185	25-65
Business	16,909	212	4.807	198	1.821	170	25-65
Science	13.439	169	3.784	156	1.436	134	25-65
High school education							
Engineering I	9,762	123	3.029	125	1.357	127	20-65
Engineering II	8.945	112	2.756	114	1.231	115	20-65
Business	7.960	100	2.427	100	1.069	100	20-65

Note: The calculations are done under the assumption that the year of birth is 1969 and that employees with highschool education start to 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.

The application of cross-section profiles for calculations of life-time salaries was criticized above. Would the life-time salary obtained from a cross-section profile differ very much from the salary obtained from a cohort-profile? Assuming constant increases β and γ_i the answer depends on β and the discount rate used. Suppose the present value of the salary stream is discounted to the first year on the labor market, say T_0, by the constant discount rate $\exp[\rho]-1$. From (3:21) the profile of discounted median salaries can be obtained

$$M(L_{Tf.}) = \exp\left[\alpha' + \beta f_0 + \sum_{i=1}^{c} (\gamma_i - \rho) D_i\right]; \tag{3:33}$$

where f_0 symbolizes a particular cohort. Observing that T_0 equals f_0+C and using (3:22) we find that the two profiles (3:30) and (3:33) are identical provided $\rho = \beta$ or which is the same, provided the discount rate equals the average increases of initial salaries. In our case this average salary increase is estimated to between six and seven per cent. When inflation in Sweden is taken into consideration, this percentage is perhaps a little low. If this is true estimates from cross-section profiles would overestimate the life-tima salary, provided the cross-section profile is not discounted too.

References

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.

Ben-Porath, Y., 1966, Lifetime Income and Economic Growth: Comment. The American Economic Review No 4:1 p.869-

Berkson, J., 1950, Are there two regressions? Journal of the American Statistical Association, Vol. 45

Blaug, M., Peston, M.H. & Ziderman, A., 1967, The Utilization of Educated Manpower in Industry. Edinburgh and London

Edgren, G., Faxén, K-O, Odhner, C-E, 1968, Lönebildning och samhällsekonomi. Rapport från expertgrupp tillsatt av SAF, LO och TCO

Eriksson, I., 1970, Ålder och Inkomst. Utkast till kap. 9 i betänkande om svenska folkets levnadsförhållande. Allmänna förlaget,Stockholm

Fase, M.M.G., 1969, An Econometric Model of Age-Income Profiles. A statistical analysis of Dutch income data 1958-1967. University of Rotterdam

Hill, T.P., 1959, An Analysis of the Distribution of Wages and Salaries in Great Britain, Econometrica, Vol. 27,3

Holm, M., 1970, Inkomstskillnader 1966 för helårs- och heltidsanställda män och kvinnor. Kapitel 5 samt bilagorna 1 och 2 i Svenska folkets inkomster, SOU 1970:34

Klevmarken, A., 1968, Lönestrukturen bland industriens tjänstemän 1957-1964. Research report from Institute of Statistics, The University of Stockholm

- ,1968, Civilingenjörernas lönebildning. CF-tidskriften (Organ för Sveriges Civilingenjörsförening) 1968, No 9

- ,1970, Konsumtionsfunktioner uppskattade från tvärsnitt och från tidsserier. Konsumtion - inkomst - åldersprofiler. Research Report from The Institute of Statistics, The University of Stockholm

- ,1971, Den svåra konsten att räkna inkomst- och löneökningar, Industriförbundets tidskrift 1971 No 4.

Lind, E., 1963, Wage Differentiation in Swedish Industry. Publications issued by The Swedish Employers' Confederation

Lydall, H., 1955, The Life Cycle in Income, Saving, and Asset Ownership, Econometrica, Vol. 23

-, 1968, The Structure of Earnings, Oxford University Press

Merrett, S., 1966, The Rate of Return to Education: A Critique. Oxford Economic Papers

Miller, H.P., 1965, Lifetime Income and Economic Growth, The American Economic Review, No 4 p. 834-

Mincer, J., 1970, The Distribution of Labor Incomes: A Survey, Journal of Economic Literature, March 1970

Morgan, J., 1962, The Anatomy of Income Distribution, The Review of Economics and Statistics, Vol. XLIV No 3

SACO, 1968, Skenbar och verklig inkomst. Sveriges Akademikers Centralorganisation

SAF, 1969, Utjämnade lönekurvor. Mimeographed, The Swedish Employers' Confederation

SAF, yearly, Löner för tjänstemän, The Swedish Employers' Confederation.