European Economic Review 4 (1973) 235–264. © North-Holland Publishing Company

THE RELATION BETWEEN INCOME, INTELLIGENCE, EDUCATION AND SOCIAL BACKGROUND

P. de WOLFF and A.R.D. van SLIJPE
University of Amsterdam, Amsterdam, The Netherlands

First version received September 1972, revised version received January 1973

In this article the authors have tried to estimate the relative importance of education, intelligence and social background for the explanation of the variations in personal income. The investigation has been carried out on Swedish data which were generously made available to them by Professor Husén of Stockholm University.

The results pointed to education as the most important factor in explaining the income variations, followed by the social class of the fathers and the level of intelligence in this order. The relations turned out to be strongly curvilinear.

Moreover, the interactions between the explanatory variables have been studied. It was found that their combined contribution is much larger than would correspond to additivity, in particular for the higher classes of the variables. They clearly reinforce each other.

1. Introduction

In 1938 the Swedish sociologist Dr. S. Hallgren started an experiment in order to identify certain formative factors in childhood and assess their relative importance for subsequent life career. He collected from all third-graders of the Malmö primary school system data on intelligence, school performance and social background. Very soon the idea came up to follow this group of about 1500 children in the age of 9–11 years as long as possible and as a matter of fact Dr. Hallgren was able to do so until his death in 1961. One of his early collaborators, Torsten Husén, now of Stockholm University continued the work and from a survey carried out in 1963 in which he was able to trace about 85% of the original group, he collected an additional set of important data, viz. on taxed income, formal education, occupation, civil status and other social and cultural factors.[1]

[1] When the first author met Prof. Husén some years ago and asked him if it would be possible to derive from his investigation an assessment of the relative importance of various

(continued on next page)

The results of his work on these data were published in 1969 in a book, entitled *Talent, opportunity and career* [1]. Husén's interest was mainly concentrated on the various aspects of the so-called life career in which the income data only played a minor role.

When our study was largely completed our attention was drawn to an article by John C. Hause [2] who had been able to use the same data. As his approach in many respects differed from ours we considered it still worthwhile to publish our findings. In the summary a brief comparison will be made between Hause's results and ours.

2. Characteristics of the empirical material

Dr. Hallgren's original investigation of 1938 was concerned with 840 boys and 710 girls. As this survey referred to all third-form pupils of the Malmö primary school system, grade repeaters of the previous year included, the less gifted children are slightly overrepresented in the sample. [2] This is also one of the reasons why not all children represented in it were born in 1928, small numbers were born in 1927 or even earlier. But as we used the data for cross-section analysis only this does not seem to have been a disturbing factor.

As we were primarily interested in explaining income levels, only those cases could be included in our investigation for which Dr. Husén and his collaborators have been able to determine the 1963 income. This proved to be possible for 692 male respondents. From this group 147 had to be eliminated for various reasons. For most of them one or more of the social factors were unknown. A small group had to be left out as the 1963 income could not be considered as an appropriate indicator of the respondent's earning power either because he was only employed

(Footnote 1 continued)

social factors such as education, intelligence and social background in the formation of income, he replied that this particular problem had not been studied but that the available data would enable one to do so. Moreover, he generously put the tape containing the basic data at our disposal and the results presented below are based on this rather unique material. We are very grateful to Prof. Husén for his much appreciated cooperation. We also want to express our appreciation for a number of constructive remarks made during the Budapest (1972) meeting of the Econometric Society where a preliminary version of this paper was presented.

[2] Less gifted children have a higher chance to repeat a form than the others; hence, the relative frequency of this group in the sample will also be somewhat higher than for the Malmö population as a whole.

part-time or unemployed during part of 1963 due to illness or other causes.

Finally, we were left with 545 male persons for whom a complete set of data was available. This still amounts to 65% of the male part of the original sample. Due to reasons mentioned, the situation for the female part of the sample was much less favourable and, therefore, it had to be left out of account.

The set of our data consisted of the following items:

(a) The 1963-taxable income, I, (made available to Dr. Husén by the fiscal authorities; in Sweden these data are not kept secret contrary to the practice in most other countries).

(b) The social class of the respondents' parents, X, as registered in 1938 by Dr. Hallgren before the group tests were administered. The classification of this item has been based mainly on the profession of the head of the family as will be clear from the labels attached to the four classes described below. Attention has also been paid to other factors, e.g. whether or not the family was receiving some form of social aid and to the income of the head of the family. With respect to the latter factor there was a large degree of overlapping between classes. Therefore, the somewhat arbitrary role has been adopted to rearrange the classification in such a way that the upper limit of a class becomes equal to the average of the next higher one and the lower limit equal to the average of the class just below. For the four classes distinguished the following results have been obtained:

Class	Lower limit	Average (in Sw. Kr. per annum)	Upper limit
1	0	1885	3058
2	1885	3058	4146
3	3058	4146	8860
4	4146	8860	∞

A more detailed description of the four classes is given below:

1. Non-skilled workers and equivalents. Incomes not exceeding 3,058 Sw.Kr., no bottom limit. Average income (after exclusion of 95 untaxed families) 1,885 Sw. Kr. All poor, and social welfare aid in one form or another received by all.

2. Skilled workers and equivalents. Incomes between 4,146 and 1,885 Sw. Kr. Average income 3,058 Sw. Kr. None destitute, practically no social welfare assistance received.

3. Small independent businessmen and employees and lower civil servants. Incomes between 8,860 and 3,058 Sw. Kr. Average income 4,146 Sw. Kr. All considered to be fully capable of supporting themselves, and no social assistance received.
4. Employers and managers in central and local government, industry, commerce, etc., and others with similar posts in most cases requiring higher education. Incomes above 4,146 Sw. Kr. Average income 8,860 Sw. Kr. All considered to be fully capable of supporting themselves, and no social assistance received.
(c) The intelligence quotient, Y, as tested in 1938 by Dr. Hallgren. The group test devised and standardized by him consisted of 4 items (opposites, missing words, a perception test and a sentence construction test). The test was administered during the spring term of 1938 and great care was given to carrying it out under as uniform conditions as possible. Dr. Hallgren used his own test score which, however, was transformed to the ordinary scale of IQ values. These have been grouped into five classes, viz.

1: 85 and below, 2: 86–92, 3: 93–107, 4: 108–115, and 5: 116 and over.

(d) The educational performance. Here, two different measures were available, viz. the number of years of schooling completed, Z, and, alternatively, the highest educational level attained, E. The first variable has been grouped into four classes:
1: less than 8 years, 2: 8–10 years, 3: 11–14 years and 4: 15 years or more.

For the second variable five consecutive levels have been distinguished. 1. Left school when mandatory school age expired. 2. Education also after expiration of mandatory school age, but not transferred to a secondary school preparing for academic training (the so-called realskola and the gymnasium). 3. Drop-outs from the secondary school system (mentioned sub. 2). 4. Graduated from the secondary school system (sub. 2) but not holding graduate degree of a university. 5. Holding graduate degree of a university.

The necessary elimination of about 1/3 of the original data probably has not had a distorting influence on the sample. At least for the two variables for which a check was possible, viz. the IQ and the social class of the parents the effect was small as can be seen from the following table.

Table 1
The distribution of the sample according to intelligence and social class.

IQ level	1938–IQ Before elimination	After elimination	Social class	1938–social class of parents Before elimination	After elimination
< 85	22%	20%	1	30%	32%
86– 92	12%	12%	2	36%	35%
93–107	35%	38%	3	20%	21%
108–115	14%	14%	4	14%	12%
> 116	17%	16%			
	100%	100%		100%	100%

In the IQ distribution the extremes at both ends have been reduced slightly more than the central groups, whereas in the social class distribution there is a weak shift to the lower classes.

Of the variables defined before I is a quantitative variable, whereas X and E, evidently, are only available as rank numbers. Y and Z occupy an intermediate position. They too had already been grouped in the data put at our disposition. However, for the intelligence variable the individual IQ were also available. Nevertheless, in the greater part of the computations both variables have been measured by the rank numbers of their classes (but cf. sec. 6).

It will be clear that there is not much point in explaining the variations in I from these rank numbers of the explanatory variables. This would imply that an increase of one unit of one of the ranks, say $\Delta X = 1$, would always lead to the same increase of I irrespective of the starting value of this variable X. As the class intervals of the explanatory variables could not be chosen but in a rather arbitrary way there is no theoretical justification for such linear relationships.

Therefore, apart from an orienting exercise described in sec. 3, a more appropriate approach has been adopted in which each variable has been represented by a series of dummy variables each corresponding to a separate class rank. E.g. X has been replaced by the set (X_1, X_2, X_3, X_4) defined in table 2.

The classes of X (and those of the other explanatory variables as well) have been ranked in increasing order of magnitude. A similar procedure has been used for Y (with 5 dummies), for Z (with 4) and for E (with 5).

Table 2
Definition of the dummy X

X	X_1	X_2	X_3	X_4
Class 1	1	0	0	0
Class 2	0	1	0	0
Class 3	0	0	1	0
Class 4	0	0	0	1

The use of dummies of the type indicated leads to linear dependence between each full set and the constant of the regression equation investigated. Therefore, all dummies with index 1 have been excluded. Consequently, the constant obtained measures the expected income for $X_1 = Y_1 = Z_1$ (resp. E_1) = 1. The regression coefficients of the dummies included refer to the differential effects due to the other values of X, Y and Z (resp. E).

The dummies can also be used to study the intercorrelation between the explanatory variables. Their number, (4 × 5 × 4 × 5=) 400, is far too large to give a complete description. Moreover, as will be shown later on, in general only the regression coefficients for the dummies with the highest indices are significantly differing from zero. Finally, the correlation between Z and E, both referring to education cannot be but high and, hence, these two variables have never been used together. Therefore, in table 3 only the intercorrelation coefficients for I, X_4, Y_5 and Z_4 are shown.[3] (For a pair of dummies the square of the coefficient of correlation is equal to the χ^2 for the corresponding four field table divided by the total sample frequency.)

Table 3
Coefficients of correlation between income and the most important dummies (sample size n = 511).

	I	X_4	Y_5	Z_4
I	1	0.487	0.276	0.553
X_4		1	0.160	0.487
Y_5			1	0.205
Z_4				1

[3] These results refer to the sub-sample from which incomes below 10,000 Sw. Kr. and above 80,000 Sw. Kr. have been eliminated, cf. sec. 4.

It follows from the table that all dummies enlisted have a rather high degree of correlation with income I, this is in particular true for the educational variable Z_4. This conclusion will be confirmed by all following results. Furthermore, the table shows a high degree of correlation between the highest social class (X_4) and the class of the longest periods of schooling (Z_4). This is no surprising result as the higher social classes in general will try to give their children the highest form of education within the range of their capacities. The correlation between X_4 and Y_5 is rather low, which confirms the fact that a high social class offers little guarantee for very intelligent offspring. The remaining coefficient, viz. the correlation between Y_4 and Z_5 (high intelligence and long period of schooling) is slightly higher. Nevertheless, it points to the fact that in pre-war Sweden even very intelligent children often did not get a chance to participate in advanced forms of education.

Table 4 considers the connection between the alternative educational variable E with income and the other social factors. It shows (for the same sub-sample as in table 3) the correlation coefficients of E_4 with I and a few dummies of X and Y. Results for E_5 have not been mentioned as there were only 6 persons (of which 4 belong to Y_5 and 2 to Y_4) with completed university education in the sub-sample which renders the correlation coefficients insignificant.

Table 4
Correlation coefficients between E_4 and I, X_2, X_3, X_4, Y_4 and Y_5.

	I	X_2	X_3	X_4	Y_4	Y_5
E_4	0.445	−0.050	0.107	0.345	0.028	0.319

As for samples of the size considered here (n = 511, cf. table 6) a correlation coefficient of about 0.08 is just significant at the 5% level. It follows that with rising social class there is a rising tendency to let the children have a secondary education of the academic type (E_4). It is a remarkable fact that the correlation between E_4 and intelligence is only significantly positive for the highest intelligence class (Y_5).

The regression analysis based on the use of dummy variables is not the only technique which can be applied to the analysis of these data. It would also have been possible to use an analysis of variance of the incomes in the various class combinations of the explanatory variables.

Apart from the close connection between the two techniques the regression analysis has been preferred as it leads directly to measures for the importance of the different explanatory variables in explaining income differentials.

3. The results obtained from the total sample

As has been mentioned in sec. 2 a first orienting experiment was made in which the three explanatory variables were each simply measured by the rank number of the classes into which they were broken down. In order to get some insight into a possibly curvilinear relation between income and the explanatory variables correlations have been computed for income itself as well as for its logarithm.

The following results have been obtained:

$$I = \underset{(60.3)}{2.49} + \underset{(27.2)}{2.16\,X} + \underset{(29.3)}{1.43\,Y} + \underset{(13.6)}{4.83\,Z} \qquad R^2 = 0.254, \tag{1}$$

$$\ln I = \underset{(2.4)}{2.22} + \underset{(34.5)}{0.060\,X} + \underset{(26.0)}{0.056\,Y} + \underset{(11.4)}{0.202\,Z} \qquad R^2 = 0.291. \tag{2}$$

In these relations I stands for the annual income in 10^3 Sw. Kr. and $\ln I$ for the natural logarithm of I (this definition has been retained throughout the whole paper). The definitions of X, Y and Z have been given in sec. 2. The figures in brackets refer to the standard errors of the corresponding coefficients expressed as a percentage of their values. The first attempt has only been carried out for the first alternative of the educational variable. In a number of more detailed calculations to be described later on, the second alternative has also been taken into account.

As was to be expected from a cross-section type of analysis of the present kind the squared correlation coefficients (R^2) were rather low, but the standard errors point to a high degree of significance of all the explanatory variables. The R^2 for the logarithmic variant is slightly higher than for the original figures. This is probably due to the fact that the influence of extremely high income is considerably reduced in the logarithmic approach.

Relation (1) easily permits an evaluation of the differences in income due to the social factors taken into account. According to this relation a man whose variables all attain their highest value ($X = 4$; $Y = 5$; $Z = 4$) may expect an income of 37,600 Sw. Kr. In the opposite case ($X = Y = Z = 1$) the expectation would be 10,900 Sw. Kr. Moreover, it is easy to see that the income difference due to the largest difference in social class ($\Delta X = 4 - 1 = 3$) is equal to $3 \times 2{,}160 = 6{,}480$ Sw. Kr. For the IQ and the education variables the corresponding differences are: $4 \times 1{,}430 = 5{,}720$ Sw. Kr. and $3 \times 4{,}830 = 14{,}490$ Sw. Kr. Therefore, already this first result shows that education is the most important (and at the same time most significant) factor to explain the differences in income. To some extent the results described above are depending on the class intervals chosen but, nevertheless, they seemed to be encouraging enough to continue the analysis with the help of the dummies introduced in sec. 2. The results for the sample as a whole are presented in table 5.

Table 5
Regressions for the total sample ($n = 545$).

$$I = \underset{(5.5)}{16.57} + \underset{(78.3)}{1.62\,X_2} + \underset{(54.5)}{2.67\,X_3} + \underset{(8.7)}{20.11\,X_4} \qquad (R^2 = 0.208) \qquad (3)$$

$$I = \underset{(7.9)}{15.63} + \underset{(80.7)}{2.50\,Y_2} + \underset{(64.2)}{2.37\,Y_3} + \underset{(25.8)}{7.51\,Y_4} + \underset{(13.5)}{13.53\,Y_5} \qquad (R^2 = 0.114) \qquad (4)$$

$$I = \underset{(4.6)}{15.81} + \underset{(39.4)}{2.99\,Z_2} + \underset{(13.8)}{10.62\,Z_3} + \underset{(8.2)}{25.61\,Z_4} \qquad (R^2 = 0.249) \qquad (5)$$

$$\begin{aligned} I = {} & \underset{(9.3)}{13.69} + \underset{(220.9)}{0.54\,X_2} + \underset{(647.4)}{0.22\,X_3} + \underset{(18.7)}{10.70\,X_4} + \underset{(68.4)}{2.61\,Y_2} + \underset{(109.9)}{1.24\,Y_3} \\ & + \underset{(44.8)}{3.95\,Y_4} + \underset{(23.6)}{7.33\,Y_5} + \underset{(87.0)}{1.36\,Z_2} + \underset{(26.5)}{5.94\,Z_3} + \underset{(16.5)}{15.16\,Z_4} \qquad (R^2 = 0.320) \qquad (6) \end{aligned}$$

$$\ln I = \underset{(1.2)}{2.74} + \underset{(54.9)}{0.08\,X_2} + \underset{(42.3)}{0.12\,X_3} + \underset{(9.2)}{0.69\,X_4} \qquad (R^2 = 0.186) \qquad (7)$$

$$\ln I = \underset{(1.6)}{2.71} + \underset{(84.5)}{0.08\,Y_2} + \underset{(57.4)}{0.09\,Y_3} + \underset{(22.0)}{0.31\,Y_4} + \underset{(12.7)}{0.51\,Y_5} \qquad (R^2 = 0.131) \qquad (8)$$

$$\ln I = \underset{(1.0)}{2.70} + \underset{(23.7)}{0.17\,Z_2} + \underset{(11.8)}{0.43\,Z_3} + \underset{(7.8)}{0.95\,Z_4} \qquad (R^2 = 0.275) \qquad (9)$$

$$\begin{aligned} \ln I = {} & \underset{(1.7)}{2.61} + \underset{(132.9)}{0.03\,X_2} + \underset{(545.9)}{0.01\,X_3} + \underset{(24.2)}{0.29\,X_4} + \underset{(73.4)}{0.09\,Y_2} + \underset{(113.3)}{0.04\,Y_3} + \underset{(37.7)}{0.17\,Y_4} \\ & + \underset{(22.9)}{0.27\,Y_5} + \underset{(35.2)}{0.12\,Z_2} + \underset{(19.5)}{0.29\,Z_3} + \underset{(14.0)}{0.63\,Z_4} \qquad (R^2 = 0.332) \qquad (10) \end{aligned}$$

(For the regressions with logarithmic income two persons with negative incomes were excluded from the population.)

They clearly show that the linear relation assumed in (1) and (2) is not valid. It is true that all regression coefficients are positive as might be expected. However, the assumption would require that the coefficients of the dummies belonging to the same variable form an increasing arithmetic series. Instead of this a considerable number is small and insignificant. In fact, in the overall eqs. (6) and (10) only the coefficients for the higher indices of all variables (X_4; Y_4, Y_5; Z_3, Z_4) are significant (at the 5% level).

With a few exceptions the same is true for eqs. (3), (4), (5), (7), (8) and (9), which show the correlations of I resp. ln I with each of the explanatory dummy sets separately. In this respect there is practically no difference between the arithmetical and the logarithmic approaches (for (6) and (10) even the R^2 values are almost identical). Moreover, all the significant coefficients show rising increases with increasing dummy indices which point to markedly curvilinear relations.

From (6) it follows that a (X_1, Y_1, Z_1) -person earns on the average 13,690 Sw. Kr. whereas a (X_4, Y_5, Z_4) -person has an average income of 46,880 Sw. Kr. The greater part of this large difference is again due to education (15,160 Sw. Kr.) and the smallest part (7,330 Sw. Kr.) to intelligence.

If we compare the equations showing the effects of each social factor separately with the overall ones we observe considerable differences for the same dummy variables. This is due to the fact, already referred to in sec. 2, that the explanatory variables themselves are intercorrelated. If one variable, e.g. education is left out of account, then part of its contribution to the explanation of income differences is automatically taken over by the remaining variables. This phenomenon pleads for a cautious interpretation even of the results of the overall equations. Moreover, the set of explanatory variables which we could take into consideration is far from complete. If it had been possible to measure such factors as industriousness, adaptability and perseverance the outcome may have been that part of the influence now ascribed to the factors X, Y and Z would be attributed to those others.

Although it follows from the preceding remarks that the one-variable equations have to be considered with still greater caution, it is interesting to note that they, nevertheless, confirm the conclusion that education is the most important explanatory factor, followed by social class and intelligence in the rear. This is valid both for the size of the corresponding values of R^2 and for the regression coefficient of the dummy variable with the highest index. Only in the logarithmic case the results are slightly different; it is true, that there too, education is most important, but the difference between the contributions of social class and intelligence are small. This result is due to the fact that the social background is the more important the higher the incomes considered and their influence is evidently reduced by using logarithms.

4. Some results for more homogeneous sub-samples

4.1. The results for the population of employees

The calculations presented in sec. 3 refer to the sample as a whole, but this group is heterogeneous. By far the largest part consist of employees (in the private and public sectors), a small fraction is independently employed (employers and workers on their own account). The possibility cannot be excluded that the relation between income and social factors for the former group differs from the latter group because variables which could not be taken into account (family relations, market position, etc.) play an important role in the latter case. The group is too small to be investigated separately but, nevertheless, it is interesting to study the effect of its elimination on the results obtained before.

A second aspect which merits further investigation is the income distribution of the sample. A large variance is a necessary condition for a study of the type made, but the sample contains a number of excessively high as well as low incomes which might distort the results. A definition of "low" or "high" must be arbitrary. For the lower cut-off point an income of 10,000 Sw. Kr. was chosen, as in view of the social conditions prevailing in Sweden in 1963, it is rather improbable that incomes below this level can be considered normal. This assumption is supported by the fact that the group (cf. table 6) contains a considerable number of independents, probably small tradesmen having an exceptionally bad year. The upper limit is fixed at 80,000 Sw. Kr. for the simple reason that the incomes above this limit showed a very high degree of dispersion. In this second case too the groups eliminated were to small for a separate study, but as in particular the higher incomes might have a distorting effect on the results (cf. the remarks made in sec. 3) it was considered worthwhile to eliminate these groups too.

The size of the various sub-samples is shown in table 6.

Table 6
Size of sub-samples.

Income in 10^3 Sw. Kr.	Employees	Independents	Total
$10 < I < 80$	474	37	511
$I < 10$	18	13	31
$I > 80$	3	0	3
Total	495	50	545

Table 7
Regressions after exclusion of the group of independents (sample size $n = 495$).

$$\begin{array}{llllll} I = 16.45 & + 1.87\,X_2 & + 3.24\,X_3 & + 18.78\,X_4 & & \\ \quad (5.6) & (68.4) & (45.2) & (9.7) & & \end{array} \qquad (R^2 = 0.187) \qquad (11)$$

$$\begin{array}{llllll} I = 16.00 & + 2.10\,Y_2 & + 1.88\,Y_3 & + 7.07\,Y_4 & + 12.04\,Y_5 & \\ \quad (7.7) & (96.9) & (81.5) & (27.3) & (15.3) & \end{array} \qquad (R^2 = 0.103) \qquad (12)$$

$$\begin{array}{llll} I = 15.97 & + 2.41\,Z_2 & + 11.00\,Z_3 & + 22.71\,Z_4 \\ \quad (4.6) & (50.0) & (13.3) & (3.6) \end{array} \qquad (R^2 = 0.229) \qquad (13)$$

$$\begin{array}{llllllll} I = 14.10 & + 0.64\,X_2 & + 0.60\,X_3 & + 10.22\,X_4 & + 2.07\,Y_2 & + 0.64\,Y_3 & + 3.95\,Y_4 \\ \quad (9.1) & (187.4) & (242.7) & (20.0) & (87.8) & (217.8) & (45.1) \\ \quad + 6.85\,Y_5 & + 0.79\,Z_2 & + 6.54\,Z_3 & + 13.37\,Z_4 & & & \\ \quad (25.6) & (151.9) & (24.3) & (18.9) & & & \end{array} \qquad (R^2 = 0.300) \qquad (14)$$

$$\begin{array}{llllll} I = 16.08 & + 7.82\,E_2 & + 4.98\,E_3 & + 13.38\,E_4 & + 32.02\,E_5 \\ \quad (4.0) & (47.1) & (32.7) & (9.8) & (15.7) \end{array} \qquad (R^2 = 0.222) \qquad (15)$$

The results obtained after the elimination of the group of independents are presented in table 7. Contrary to expectation the results, which are only given for the non-logarithmic case, do not differ very much from the corresponding ones in table 5. The same regression coefficients are significant and their differences are very small. Remarkably enough, the correlation coefficients are somewhat lower than in table 5. The constant in the overall equation is slightly larger which, in accordance with the lower correlation coefficients, points to a lower contribution of the social factors to the explanation of the income variations.

For the same sub-sample the correlation has been computed between I and the alternative educational variable E described in sec. 2 (cf. eq. (15)). The indices of the set of dummies belonging to this variable correspond to the rank numbers of the levels of education obtained. In this case all the regression coefficients are significant at the 5% level. The most striking result is the very high contribution of a university degree (about 33,000 Sw. Kr. or nearly 20,000 Sw. Kr. above the next lower level). It is also interesting to note that the coefficient of E_2 is larger than that of E_3. This might be expected. The group corresponding to E_2 consists of the persons who after leaving primary school did not go on to secondary schools preparing for academic training but it is probably a positive selection of the primary school leavers as several of them have been able to complete certain types of tertiary (non-university) training. The E_3 category, however, consists of the drop-outs from the academic type of secondary school and, therefore, probably forms a negative selection. The fact that the coefficient of E_4 surpasses both the coefficients of E_2 and E_3 shows that a complete secondary education of the academic type is to be considered as an advantage in comparison to the other two types.

An overall equation corresponding to (14) has not been computed; this has only been done for the sub-sample obtained by eliminating the excessive incomes.

4.2. *The results for the population of employees after exclusion of the excessive incomes*

In a second experiment the sub-sample studied was obtained by excluding not only the independent income earners but also the group with excessive incomes described in sec. 4.1. The outcome was very encouraging as the R^2-value for the overal equation rose to 0.472 which is a rather high figure for a cross-section analysis of a sample with 474 elements (R = 0.69; whereas its standard deviation is only 0.024; cf. table 8). However, a more important conclusion is that the results obtained in sec. 3 are not caused by a few extreme values but that, on the contrary, these extremes have a disturbing influence on the results for incomes in the normal range. It supports the conjecture that excessive incomes are largely determined by factors which could not be taken into account. A possible explanation for the very low incomes has already been mentioned in sec. 4.1. As to the very high incomes, one can think of the incomes and the fortunes of the parents as distinct from their social position.

The significant regression coefficients remain of the same order of magnitude with a few important differences. The most striking one is the reduction of the coefficient of X_4 by nearly 50%. It may be due to the fact that the highest social class is very heterogeneous. On the one

Table 8
Regressions after exclusion of the group of independents and a small number of excessively low and high incomes (below 10,000 resp. above 80,000 Sw. Kr.; sample size $n = 474$).

$$I = \underset{(5.5)}{14.30} + \underset{(81.2)}{0.93\,X_2} + \underset{(163.5)}{0.54\,X_3} + \underset{(21.6)}{5.77\,X_4} + \underset{(41.5)}{2.71\,Y_2} + \underset{(71.4)}{1.20\,Y_3} + \underset{(26.6)}{4.08\,Y_4}$$
$$+ \underset{(23.6)}{4.52\,Y_5} + \underset{(50.0)}{1.49\,Z_2} + \underset{(16.2)}{6.01\,Z_3} + \underset{(9.0)}{17.00\,Z_4} \qquad (R^2 = 0.472) \qquad (16)$$

$$\ln I = \underset{(1.2)}{2.66} + \underset{(56.0)}{0.06\,X_2} + \underset{(143.2)}{0.03\,X_3} + \underset{(26.9)}{0.19\,X_4} + \underset{(37.5)}{0.12\,Y_2} + \underset{(53.6)}{0.07\,Y_3} + \underset{(26.4)}{0.17\,Y_4}$$
$$+ \underset{(20.7)}{0.21\,Y_5} + \underset{(32.9)}{0.09\,Z_2} + \underset{(13.3)}{0.30\,Z_3} + \underset{(9.9)}{0.64\,Z_4} \qquad (R^2 = 0.447) \qquad (17)$$

$$I = \underset{(2.1)}{17.50} + \underset{(14.8)}{7.65\,X_4} + \underset{(20.5)}{4.26\,Y_5} + \underset{(10.4)}{14.20\,Z_4} \qquad (R^2 = 0.395) \qquad (18)$$

$$\ln I = \underset{(0.6)}{2.83} + \underset{(17.0)}{0.28\,X_4} + \underset{(17.8)}{0.21\,Y_5} + \underset{(12.9)}{0.49\,Z_4} \qquad (R^2 = 0.335) \qquad (19)$$

$$I = \underset{(2.5)}{16.63} + \underset{(31.8)}{7.27\,E_2} + \underset{(22.3)}{4.63\,E_3} + \underset{(7.5)}{11.13\,E_4} + \underset{(10.0)}{32.37\,E_5} \qquad (R^2 = 0.364) \qquad (20)$$

$$I = \underset{(5.3)}{15.09} + \underset{(238.8)}{0.33\,X_2} + \underset{(9082.2)}{0.01\,X_3} + \underset{(17.3)}{7.60\,X_4} + \underset{(47.6)}{2.44\,Y_2} + \underset{(166.8)}{0.54\,Y_3} + \underset{(32.4)}{3.55\,Y_4}$$
$$+ \underset{(33.4)}{3.43\,Y_5} + \underset{(116.2)}{2.03\,E_2} + \underset{(28.0)}{3.70\,E_3} + \underset{(12.4)}{7.85\,E_4} + \underset{(14.7)}{22.83\,E_5} \qquad (R^2 = 0.436) \qquad (21)$$

hand this class consists of families where the father had a reasonable income, e.g. due to an academic training which enabled him to allow his children to attain the highest educational levels corresponding to their abilities. On the other hand it contains families where the father occupied a high position in business which enabled him not only to offer his children the same advantages as the parents belonging to the former group but in addition the possibility to attain jobs in the highest income brackets.

By eliminating precisely these high incomes the weight attributed to X_4 is reduced. The fact that the importance of a long schooling period is increased (coefficient of Z_4) points in the same direction. It is less clear why the weight of high intelligence has also been diminished but the difference is less significant than for the other two variables.

Looking at the overall equation (16), it is clear, that within the income range 10,000–80,000 Sw. Kr. the order of the variables according to their importance for the explanation of income differentials is the same as in the previous cases, viz. education, social background, intelligence. But the relative importance of education has increased considerably. In this case even the coefficient of the highest educational class but one (Z_3) surpasses the coefficient of the highest social class (X_4). This effect is still more pronounced with the logarithmic approach (17). As this one tends to reduce the weight of the higher incomes it clearly shows that the lower the social class the more important becomes the length of the schooling period for the determination of the income. This is to be seen as a continuation of the effect caused by the elimination of the highest incomes.

From (18) and (19) one can see that the three dummies with the highest indices (X_4, Y_5 and Z_4) contribute the essential part of the explanation. For the arithmetical approach (18) these three variables explain nearly 40% ($R^2 = 0.395$) of the total variation of I. In the overall equation (16) this share has been increased by not more than about 8% ($R^2 = 0.472$).

4.3. The results for the sample obtained by eliminating only the excessive incomes

In sec. 4.1 and 4.2 the results described have been obtained by consecutively eliminating the independent workers and the excessive incomes. It was shown that the first step had very little effect but that the second led to a considerably higher R^2 and also to a marked change in some of the regression coefficients. Therefore, a fourth alternative has been considered, viz. eliminating (from the total sample) the excessive incomes only. As might be expected the results for this sub-sample ($n = 511$, cf. table 6) do not deviate from those described in sec. 4.2 (table 8). The correlation coefficients are slightly lower than in sec. 4.2 but the regression coefficients show very small deviations. This check proves again that, insofar as the small group of independents permits any conclusion at all, its behaviour does not deviate significantly from that of the population of employees, neither in the total sample nor in the case of the truncated income distribution.

The results obtained so far are illustrated in fig. 1 showing for the non-logarithmic case the regression coefficients per set of dummies corresponding to each of the variables X, Y and Z. Coefficients significant at the 5% level have been marked by a small square, the others by a dot. The figure shows the curvilinear relationship between income and each of the explanatory variables.

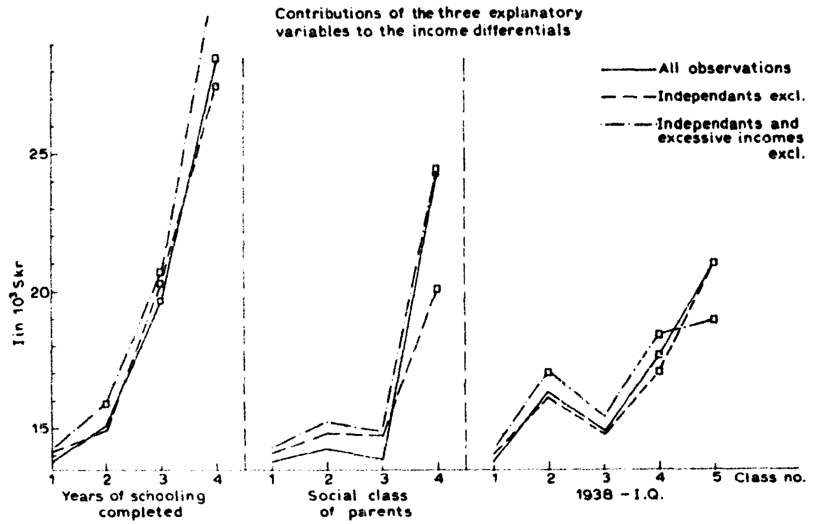

Fig. 1. Corrigendum: In the middle graph the two broken lines joining 3 and 4 must be interchanged.

5. The effect of interactions

As a final stage of the investigation an effort has been made to measure the interactions between the various explanatory variables. This could not be done without further condensation of the material. The number of classes into which the data have been broken down in sec. 4 and 5, viz. 4 × 5 × 4 = 80, was far too large to be feasible. As follows from the preceding results, the effect of the lower classes in general was very moderate and their regression coefficients often were insignificant. Therefore, the disadvantage of a coarser grouping did not seem to be too great.

The following groups have been chosen:

$$
\begin{array}{lll}
X_a = X_1 + X_2 + X_3 & X_b = X_4 & \\
Y_a = Y_1 + Y_2 + Y_3 & Y_b = Y_4 + Y_5 & \\
Z_a = Z_1 + Z_2 & Z_b = Z_3 & Z_c = Z_4 \\
E_a = E_1 & E_b = E_2 + E_3 & E_c = E_4 + E_5 .
\end{array}
$$

It would have been interesting to keep the two classes E_4 and E_5 separated in order to measure the effect of a university degree. But as has been remarked already the size of class E_5 was too small for this purpose.

The grouping described above leads to $2 \times 2 \times 3 = 12$ classes for each of the sets (X, Y, Z) and (X, Y, E). Consequently only 11 dummies can be used if the constant is maintained in the regression. This constant refers again to the "zero class", viz. X_a, Y_a, Z_a (resp. X_a, Y_a, E_a) and the regression coefficients to the differences between this class and the other one.

The results have only been computed for the smallest of the subsamples considered, viz. excluding independents and earners of excessive incomes. The results are given in table 9. In order to facilitate the comparison of the coefficients they have been grouped in a three-dimensional table. In this way it is not only possible immediately to read off the effect due to a particular combination of dummies (X_i, Y_j, Z_k) or (X_i, Y_j, E_k) with $i = a, b; j = a, b$ and $k = a, b, c$ but also to compare combinations of dummies in which only one factor has been changed (moving vertically for change in X, diagonally for changes in Y and horizontally for changes in Z in (22) and (23) and for E in (24) and (25)).

Moreover, the coefficients for the logarithmic cases have been represented graphically in fig. 2, where also the size of the cells corresponding to the combination of X, Y, Z resp. X, Y, E has been indicated. From these figures a serious disadvantage of the high degree of aggrega-

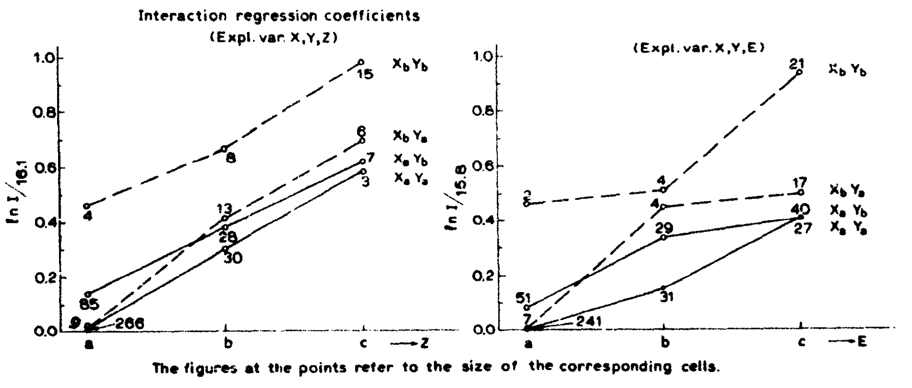

Fig. 2. Here again, as has been explained on p. 240, there is a linear dependence between the full set of dummies and the constant of the regression equation.

Table 9

Regression coefficients for interactions between explanatory variables (population exclusive of independents and excessive incomes; sample size n = 474).

I. non-logarithmic case, explanatory variables X, Y, Z (22)

X \ Z		a	b	c
a	Y_a	0	5.73 (21.9)	14.03 (27.0)
	Y_b	2.62 (31.0)	8.54 (15.2)	14.65 (17.1)
b	Y_a	0.03 (7063)	9.21 (20.1)	16.86 (16.0)
	Y_b	11.36 (28.9)	16.36 (14.3)	29.16 (5.9)

constant 16.64 (2.4) R^2 = 0.484

II. logarithmic case, explanatory variables X, Y, Z (23)

X \ Z		a	b	c
a	Y_a	0	0.30 (17.9)	0.58 (27.5)
	Y_b	0.14 (24.1)	0.40 (13.8)	0.63 (16.8)
b	Y_a	0.01 (1022)	0.41 (19.0)	0.70 (16.2)
	Y_b	0.47 (29.8)	0.67 (14.8)	0.98 (7.4)

constant 2.78 (0.60) R^2 = 0.429

III. non-logarithmic case, explanatory X, Y, E (24)

X \ E		a	b	c
a	Y_a	0	2.74 (46.4)	8.19 (16.5)
	Y_b	1.36 (75.5)	6.71 (19.5)	8.45 (13.5)
	Y_a	−0.04 (−6063)	8.29 (37.6)	11.44 (14.6)
	Y_b	11.17 (42.3)	11.67 (28.7)	26.34 (6.8)

constant 16.33 (2.6) R^2 = 0.463

IV. logarithmic case, explanatory variables X, Y, E (25)

X \ E		a	b	c
a	Y_a	0	0.15 (34.4)	0.40 (14.0)
	Y_b	0.08 (51.8)	0.33 (16.5)	0.40 (11.8)
b	Y_a	0.01 (1406)	0.45 (30.7)	0.49 (14.2)
	Y_b	0.46 (42.8)	0.51 (27.3)	0.93 (6.8)

constant 2.76 (0.70) R^2 = 0.430

The figures in brackets as usual refer to the percentage standard errors of the corresponding regression coefficients.

tion immediately can be seen. Somewhat over 50% of the whole sample is concentrated in the lowest cells (X_a, Y_a, Z_a resp. X_a, Y_a, E_a). Several of the other cells are only weakly occupied. Another disadvantage which cannot be seen from the figures presented is that due to the aggregation the X_a- and Y_a-values in general refer to larger groups than the corresponding b-values. Therefore, the comparison of e.g. X_a, Y_a, Z_b with X_a, Y_a, Z_c suggests that only the value of Z is different for these two cells, but in fact the actual composition of the X_a, Y_a combination might also have been changed. Nevertheless, the fact that all the regression coefficients (apart from X_b, Y_a, Z_a resp. X_b, Y_a, E_a which are almost equal to zero) have moderate standard errors and that their mutual relationship shows a quite feasible pattern, in our opinion, yields sufficient justification to draw a few conclusions.

As in the previous cases the logarithmic and non-logarithmic variants do not show systematic differences. We restrict our comments to the logarithmic ones as their differences, in particular for the combination (X, Y, Z), are more constant than for the other two. That is also the reason why these variants have been chosen for the graphical representation.

It is a striking fact that the slopes of the lines connecting points with different Z-values for each of the four combinations of social class and intelligence show little variation. For the lowest combination X_a, Y_a the difference of one Z-class is equal to 0.29 which corresponds to an increase of 33% (ln 1.33 = 0.29) with respect to the average income of the preceding class (the average income of the class X_a, Y_a, Z_a is 16,000 Sw. Kr.; ln 16.1 = 2.78, the constant of the regression equation). This shows that even for a rather low social class and a moderate level of intelligence it "pays" to be schooled for a long period. However, the class X_a, Y_a, Z_c probably consists of plus variants with respect to perseverance and zeal, variables which, as has been said before, could not be taken into account; for it would not be sensible to conclude that many years of schooling alone would produce the result obtained.

For the combination X_b, Y_a the effect of schooling is still somewhat larger, viz. on the average 0.35 or 42%. Here again one of the cells, viz. X_b, Y_a, Z_a is a rather remarkable one. It probably consists mainly of minus variants. In spite of the high social class of the parents together with low intelligence they have not had more than 10 years of schooling, which is probably due to lack of interest or ability. The groups X_a, Y_b and X_b, Y_a do not differ very much with respect to their reaction to Z. Both are clearly above X_a, Y_a. The higher social class leads to a slightly increasing difference, the higher intelligence to a slightly decreasing one. On the average the difference is 0.11 or 12%.

The figures show that a relatively small number of highly intelligent children from the lower social class got a chance to follow advanced education. Those who had this chance could exploit their ability less efficiently than the intelligent children of the higher classes and those who did not have this chance were also less well off than their equivalents from the higher classes. However, as could be expected the combination X_b, Y_b is a very favourable one. Here the difference with respect to X_a, Y_a on the average is 0.41 or 51% and hence roughly 90% more than what would result if the effects of social class and intelligence were additive.

The pattern for (the logarithmic variant) of the combination X, Y, E is less regular than for the previous one. We find again the practical absence of any effect of the social class for the combination of low intelligence and low level of education (Y_a, E_a) but the small difference between (X_a, Y_b) and (X_a, Y_a) is also reduced to zero if Z_c is replaced by E_c.

The average slope of the E lines for the combinations X_a, Y_b and X_a, Y_a is only 0.18 or 20% which is much less than in the corresponding case for Z (0.27 or 31%). This is due to the fact that the class distributions for the two variables differ considerably. This is immediately clear from the occupancy of the cells which for E_c is much higher than for Z_c. Fifteen years of schooling or more (Z_c) corresponds in general to a higher educational level than a completed secondary school of the academic type or more.

The E_a- and E_b-levels for the combinations X_b, Y_a and X_b, Y_b do not offer much room for conclusions as the number of respondents is very small. However, it is clear that a very high proportion of the higher social class families find a way to let their children complete a secondary school of the academic type (69% of the sample) irrespective of their intelligence. Finally, in this case too there is a very favourable effect of the combination X_b, Y_b in the case of E_c. Here the difference with respect to either X_a, Y_a or X_a, Y_b amounts to 0.53 or 70%.

6. Some supplementary results

Although the main results of the investigation have been presented in the preceding sections the available set of data permitted us to study a few variants and additions to be discussed here.

(1) So far the intelligence has been measured by the ranks of the five classes into which the IQ figures have been grouped. But as is mentioned in sec. 2 the individual data were also available and, therefore, a set of regressions has been carried out in which the variable Y has been replaced by the variable IQ, representing the original measurements reduced to the customary IQ scale. In this way the loss of information due to the classification can be avoided but there is a disadvantage connected with this procedure. It automatically leads to the assumption of a linear relation between I (or $\ln I$) and IQ and from the preceding results it is known that this assumption does not hold (cf. fig. 1). It would, of course, have been possible to use a function of IQ the general shape of which could be based upon the results obtained from Y but this has not been attempted.

(2) As an additional variable the civil status of the respondents has been introduced as a dummy variable CS, assuming the value 1 for bachelors and 2 for all other cases.

(3) The educational performance is neither satisfactorily measured by Z nor by E. The same number of years of schooling Z may lead to considerably differing levels of E; on the other hand, given the level of E a person's ability will no doubt be dependent on the number of years of schooling needed to attain this level. Therefore, a new measure S (schooling) has been worked out which takes into account both duration and level of education. It is characterised by 7 classes as follows:

S_1: Left school when mandatory school age expired according to School Act § 47 (roughly equivalent to 6 years of primary school completed),

S_2: Left school when mandatory school age expired according to School Act § 48 (mainly including cases which did not complete the 6th form and, hence, a negative selection as compared to S_1),

S_3: Education after expiration of mandatory school age but not graduated from the secondary school system, 10 or less years of schooling,

S_4: As S_3 but with 11 or more years of schooling,

S_5 : Graduated from realskola [4] but no university degree,
S_6 : Graduated from gymnasium [4] but no university degree,
S_7 : Holding a university degree.

The relation between the new variable S and the previously used ones Z and E is shown below

$$
\begin{aligned}
S_1 + S_2 &= E_1 \\
S_3 &= (E_2 + E_3) \wedge (Z_1 + Z_2) \\
S_4 &= (E_2 + E_3) \wedge (Z_3 + Z_4) \\
S_5 + S_6 &= E_4 \\
S_7 &= E_5 .
\end{aligned}
$$

The population for which the regressions based on the new variables have been carried out is the same as used in 4.2 and 5 ($n = 474$), (the sub-sample of employees after exclusion of excessive incomes). The distribution of the population over the various dummy classes is shown in table 10.

Table 10
Distribution of the sub-sample over the dummy classes.

Class \ Variable	X	S	CS
1	153	275	38
2	164	26	436
3	102	48	
4	55	20	
5		89	
6		11	
7		5	
Total	474	474	474

The (inter)correlation coefficients between I, X_4, IQ, S_5 and CS are shown in table 11.

Finally, the most interesting results of the regression analysis are to be found in table 12. From this table the following conclusions may be drawn.

[4] The final certificates of both the realskola and the gymnasium qualify for certain types of academic studies. However, the curriculum of the second type (i.e. due to the study of classical languages in addition to the subjects taught at the realskola) is more difficult.

Table 11
Correlation coefficients between I and some of the explanatory variables (sample size $n = 474$).

	I	X_4	IQ	S_5	CS
I	1	0.476	0.351	0.339	0.087
X_4		1	0.157	0.264	0.034
IQ			1	0.326	0.086
S_5				1	0.003
CS					1

Table 12
Regression obtained with X, S, IQ and CS (employees, excessive incomes excluded; sample size $n = 474$).

$$\begin{array}{llllll} I = 15.11 & -\ 1.69\,S_2 & +\ 3.78\,S_3 & +\ 7.51\,S_4 & +\ 9.57\,S_5 & +\ 22.16\,S_6 \\ \quad (7.7) & (-84.3) & (28.6) & (21.3) & (8.8) & (9.6) \\ & +\ 32.07\,S_7 & +\ 1.81\,CS_2 & & & \\ & (9.7) & (64.8) & & & \end{array} \qquad R^2 = 0.415$$

$$\begin{array}{llllll} I = 14.78 & +\ 42\,X_2 & +\ 0.11\,X_3 & +\ 6.04\,X_4 & -\ 1.42\,S_2 & +\ 3.59\,S_3 \\ \quad (8.1) & (187.2) & (833) & (20.9) & (-98.6) & (29.7) \\ & +\ 6.21\,S_4 & +\ 8.01\,S_5 & +\ 18.6\,S_6 & +\ 26.4\,S_7 & +\ 1.77\,CS_2 \\ & (25.7) & (11.3) & (11.8) & (12.3) & (64.8) \end{array} \qquad R^2 = 0.447$$

$$\begin{array}{llllll} I = 8.25 & +\ 0.43\,X_2 & -\ 0.07\,X_3 & +\ 6.18\,X_4 & +\ 0.070\,\mathrm{IQ}\ + & \\ \quad (31.1) & (180.8) & (-1229) & (20.3) & (34.7) & \\ & -\ 0.06\,S_2 & +\ 2.95\,S_3 & +\ 5.41\,S_4 & +\ 7.03\,S_5 & +\ 17.55\,S_6\ + \\ & (-2441) & (36.7) & (29.8) & (13.7) & (12.6) \\ & +\ 24.74\,S_7 & +\ 1.55\,CS_2 & & & \\ & (13.2) & (73.7) & & & \end{array} \qquad R^2 = 0.457$$

$$\begin{array}{llllll} \ln I = 2.37 & +\ 0.03\,X_2 & +\ 0.00\,X_3 & +\ 0.21\,X_4 & +\ 0.003\,\mathrm{IQ} & \\ \quad (4.4) & (98.7) & (1851) & (24.4) & (31.7) & \\ & -\ 0.02\,S_2 & +\ 0.16\,S_3 & +\ 0.27\,S_4 & +\ 0.33\,S_5 & +\ 0.63\,S_6 \\ & (-248.6) & (27.9) & (24.2) & (12.1) & (14.4) \\ & +\ 0.78\,S_7 & +\ 0.10\,CS_2 & & & \\ & (17.1) & (49.1) & & & \end{array} \qquad R^2 = 0.433$$

(a) Just as in the previous cases a significant regression coefficient is only obtained for the highest class of the social variable (X_4). Moreover, its value differs very little from the results obtained before (cf. table 8). The corresponding effect, therefore, seems to be rather stable.

(b) The regression coefficient for IQ (in those cases where it has been included) is hardly significant at the 5% level, in spite of the refined measure used. Moreover, its effect turns out to be much less than that obtained when using the corresponding dummies. The estimated income differential for an increase of IQ (ceteris paribus) from 80 (average of class 1) to 120 (average of class 5) amounts to $(120-80) \times 0.07 \times 10^3 = 2{,}800$ Sw. Kr. in comparison to 4,080 Sw. Kr. According to table 8, eq. (16). This result must be due to the non-linear relation between income and intelligence (cf. fig. 1). Therefore, the conclusion may be drawn that the substitution of the set $Y_1 \ldots Y_5$ by IQ leads to inferior results and is only justifiable if at the same time IQ is replaced by a suitable curvilinear function f(IQ) (with one or more independent parameters).

(c) The introduction of S, however, proves to be successful. S_2 shows a small and not significant difference with respect to S_1 but in all four cases mentioned its sign is in accordance with the fact that the respondents from this class form a negative selection out of the total population of primary school leavers (without further education). Also the redistribution of the former classes E_2 and E_3 over S_3 and S_4 has been favourable. Now there appears a significant difference between the classes and the correct sign. Clearly, within the whole group $(E_2 + E_3)$ the number of years of schooling has a greater discriminatory power with respect to income than the fact that one of its members got some further education but outside the academic type of secondary school (realskola and gymnasium) or was a drop-out from a school of this type.

The split of E_4 into S_5 and S_6 (the two different types of academic secondary school) again led to an interesting result as the regression coefficient for the latter class turned out to be much higher than for the former. Obviously, the final diploma of the gymnasium (without further academic training) offers considerably better income opportunities than the corresponding one of the realskola. Finally, in view of the fact that the two classes S_7 and E_5 are identical, it is satisfactory that the coefficients of S_7 in eqs. 2 and 3 of table 12 do not differ very much from the coefficient of E_5 in the last equation of table 8.

(d) It is usually assumed that the marital status has a positive influence on income and different explanations are offered. Supporting a family increases the needs and, consequently, may lead to greater efforts to obtain a higher income. Energy, sociability and adaptability are appreciated both by society and by women and, hence, at the same time may lead to higher incomes and to higher chances of marriage. The causality may also be the other way round as men probably are more attractive marriage prospects the higher their income.

Our results confirm to a certain extent the assumption formulated above. In all cases reported in table 12 the coefficient of CS_2 is positive, but it is nowhere significant at the 5% level. Neither is this the case for the rather low intercorrelation between I and CS (0.08). But the size of the effect in the arithmetical approach is in all cases quite considerable, it varies from 1500 to 1800 Sw. Kr. As an average this figure will overestimate the effect for the lower incomes. This is in accordance with the logarithmic result; the coefficient 0.10 corresponds to roughly 10% which for the lower incomes (around 10,000 Sw. Kr. in 1963) is considerably lower.

In addition to the three variants mentioned in the beginning of this section still a fourth one has been studied. Apart from the IQ of the respondents measured by Hallgren at the beginning of the investigation in 1938 at an age of around 10, the IQ is measured again some eight years later of those who were recruited for military service. It is well known that it is very difficult to design a culture-free intelligence test. Moreover, it is rather probable that as the age at which the test is administered is increasing its results will to a growing extent be influenced by other factors than intelligence, such as social background and education. Therefore, an attempt has been made to measure the effect of substituting the first IQ measurement by the later one (indicated by MIQ).

Obviously, this could only be done for the group for which both figures were available and as, for reasons described before, in this case too the excessive incomes have been left out of account the sample size was reduced to 433. For this group a regression analysis was made in which I was explained from the dummy sets for X and for S, the dummy for CS and alternatively IQ and MIQ. The same was done for $\ln I$ so that in all four equations have been obtained which are given in table 13.

The results confirm the expectations. Although the differences are not very significant, both for I and $\ln I$ the coefficient of IQ is considerably lower than the corresponding one for MIQ. On the other hand practically all the coefficients for the sets of X and S become lower in the regressions with MIQ. Only the coefficients of CS remain un-

Table 13
Regressions obtained with X, S, CS and alternatively IQ and MIQ (employees and independents for which both IQ and MIQ were available, excessive incomes excluded; sample size $n = 433$).

$$I = \underset{(33.7)}{8.69} + \underset{(220.6)}{0.39\, X_2} + \underset{(607.6)}{0.17\, X_3} + \underset{(23.8)}{6.03\, X_4} + \underset{(45.3)}{0.063\, \mathrm{IQ}} - \underset{(-922.3)}{0.19\, S_2}$$
$$+ \underset{(45.5)}{2.55\, S_3} + \underset{(35.7)}{5.59\, S_4} + \underset{(16.5)}{6.52\, S_5} + \underset{(19.4)}{13.40\, S_6} + \underset{(13.6)}{25.90\, S_7} + \underset{(59.8)}{2.08\, CS_2} \qquad R^2 = 0.394$$

$$I = \underset{(41.9)}{7.51} + \underset{(276.0)}{0.31\, X_2} + \underset{(436.3)}{0.23\, X_3} + \underset{(25.9)}{5.55\, X_4} + \underset{(41.1)}{0.076\, \mathrm{MIQ}} + \underset{(8204)}{0.02\, S_2}$$
$$+ \underset{(50.5)}{2.33\, S_3} + \underset{(40.5)}{5.02\, S_4} + \underset{(20.4)}{5.84\, S_5} + \underset{(21.0)}{12.62\, S_6} + \underset{(13.8)}{25.57\, S_7} + \underset{(59.1)}{2.10\, CS_2} \qquad R^2 = 0.395$$

$$\ln I = \underset{(5.0)}{2.40} + \underset{(194.0)}{0.02\, X_2} + \underset{(400.0)}{0.01\, X_3} + \underset{(28.9)}{0.20\, X_4} + \underset{(40.6)}{0.003\, \mathrm{IQ}} - \underset{(-238.0)}{0.03\, S_2}$$
$$+ \underset{(34.1)}{0.14\, S_3} + \underset{(28.5)}{0.29\, S_4} + \underset{(14.3)}{0.31\, S_5} + \underset{(20.5)}{0.52\, S_6} + \underset{(17.7)}{0.81\, S_7} + \underset{(44.6)}{0.11\, CS_2} \qquad R^2 = 0.375$$

$$\ln I = \underset{(5.5)}{2.33} + \underset{(242.6)}{0.01\, X_2} + \underset{(314.3)}{0.01\, X_3} + \underset{(32.4)}{0.18\, X_4} + \underset{(35.6)}{0.004\, \mathrm{MIQ}} - \underset{(-394.0)}{0.02\, S_2}$$
$$+ \underset{(37.6)}{0.13\, S_3} + \underset{(32.1)}{0.26\, S_4} + \underset{(17.8)}{0.27\, S_5} + \underset{(22.7)}{0.48\, S_6} + \underset{(18.1)}{0.79\, S_7} + \underset{(44.3)}{0.11\, CS_2} \qquad R^2 = 0.378$$

changed. Hence, part of the variance of I (or $\ln I$) explained by X and S when intelligence is measured by IQ is transferred to MIQ when this variable is substituted for IQ.

The same tendency can be derived from the intercorrelation coefficients of the two intelligence variables with I and some of the explanatory factors. In table 14 these figures are presented.

Table 14
Intercorrelation coefficients between intelligence, social background, education, civil status and income.

	X_4	S_5	CS_2	I	MIQ
IQ	0.129	0.365	0.105	0.328	0.717
MIQ	0.311	0.515	0.085	0.434	

It is clear that there is still a high correlation between IQ and MIQ, but on the other hand assuming that the standard deviations of the two variables are practically equal (which is very plausible) IQ can only explain about 70% of the variations of MIQ, At the same time it is shown that the correlation between X_4, S_5 and I is very significantly higher for MIQ than for IQ. Only for CS_2 the result is almost the same, even slightly lower.

7. Summary

In the preceding analysis, in particular in sections 3 and 4 where no attention has been given to interactions, all results pointed to education as the most important factor to explain income differentials, followed by the social class of the parents and then level of intelligence.

For the sample as a whole the non-logarithmic R^2 for all dummy variables corresponding to X, Y and Z was equal to 0.320. Hence 32% of the total variance of I could be explained from variations in the social factors. The contribution of each of them to this result is given in table 15.

Table 15
Contribution of the social factors to the explained part of the variance of I (total sample; $n = 545$).

Social class of parents (X)	Intelligence level (Y)	Education, years of schooling (Z)	Total (R^2)
11.3%	6.1%	14.6%	0.320
35.3%	19.1%	45.6%	R^2 = 100%

Not less than 45.6% of the total part of the variance explained is due to education. The results are still more outspoken for the truncated sample where independents and earners of excessive incomes have been excluded. R^2 rises to 0.472 and the percentage due to education to 66.4 (cf. table 16).

Table 16
Contribution of the social factors to the explained part of the variation of I (reduced sample; $n = 474$).

Social class of parents (X)	Intelligence level (Y)	Education, years of schooling (Z)	Total (R^2)
9.4%	6.5%	31.4%	0.472
19.9%	13.7%	66.4%	R^2 = 100%

The latter result is the more important as it refers to incomes of more normal character.

The two factors which may be considered as providing pure rents (social class and intelligence) are much less important than education which requires substantial sacrifices in costs and incomes foregone and, hence, an important part of the income differentials may be seen as a compensation for these sacrifices.

Another conclusion from this analysis is the fact that the relation between income and the explanatory variables is rather curvilinear in the sense that the lower classes of each of these variables contribute very little, whereas the higher two or three yield increasing returns.

The important conclusion from the interaction analysis in sec. 5 is that intelligence and social class are reinforcing each other. Their combined contribution in particular for the higher levels of education is much larger than would correspond to additivity of the effects.

In sec. 6 it was shown that civil status had a considerable and significant effect on income. Bachelors on the average earned 10% less than those who, all other factors being the same, were or had been married.

The introduction of a new variable S to measure education, being a combination of the level of education obtained and the number of years of schooling, proved to be fruitful. In particular, as it showed that the final diploma of a gymnasium offered better income prospects than the corresponding one of a realskola. However, as may be derived from table 13, the percentage of the income variance explained by S is hardly higher than years of schooling as an explanatory variable. Instead of the 66.4% of table 16 a percentage of 67.6 is found. Finally, the impression is gained that the IQ measured at the age of 9 is considerably more culture-free than when measured at entrance into military service.

The introduction mentioned that John C. Hause almost simultaneously made use of the same material as we did. In a sense his investigation was even more extensive as he also had the respondents 1968 income at his disposal. His results are only partly comparable with ours. In the first place Hause used the Swedish results to compare them with the outcome of three more or less analogous American results and, therefore, his analysis is much less elaborate than ours. Roughly the same explanatory variables are used in his and in our study, though Hause does not make use of the variable Z (years of schooling). Furthermore, he restricts himself to a study of the direct effects of the explanatory variables, in particular, intelligence, and no attempt has been made to study interactions.

Finally, a comparison is made difficult by the fact that Hause computed regressions of the logarithm on income on the social variables for each educational level separately. However, he finds that in these regressions the coefficient for the IQ (assuming a linear relation between IQ and ln I which according to our findings seems to be less satisfactory) in general increases with the level of education which again points to a certain curvilinearity of the type which we found.

We end this summary as far as our own results are concerned with a warning. Irrespective of the educational variable used a little less than 50% of the total variance of I could be explained. Satisfactory as this may seem it should not be forgotten that more than half of the variance has to be attributed to factors which could not be seized. Moreover, our result is probably favoured by the fact that the incomes were taken at the age of about 35, when particularly the higher income earners are not yet at the top of their careers. In table 17 some Dutch income data are presented which show that for all levels of education incomes as well as their *relative* coefficients of variation tend to rise with increasing age.

Table 17
Income distribution depending on age and level of education (Dutch data)*.

Level of education		Private sector			Public sector		
		Age in years					
		25–30	35–40	50–65	25–30	35–40	50–65
Primary school	(a)	8.7	10.9	11.7	9.4	10.5	11.4
only	(b)	21	25	35	16	14	34
Secondary school	(a)	10.0	14.7	17.2	10.8	14.2	18.1
(acad. type)	(b)	23	31	51	21	28	38
University degree	(a)	18.1	27.3	40.1	16.1	23.4	32.6
	(b)	18	30	50	12	22	35

(a) Median 1967-incomes (in 1000 guilders/annum).
(b) Quartile distance in % of median income.
* Sociale Maandstatistiek C.B.S. Aug. 1967, Nov. 1967.

Therefore, if it would be possible for Professor Husén to survey his sample again, say in 1983, an entirely different part of the variation of the then prevailing incomes could be explained by the three social factors investigated. It is difficult to guess whether this part would be greater or smaller than in 1963.

References

[1] Torsten Husén, Talent, opportunity and career, Almquist & Wicksell, Stockholm, 1969.
[2] John C. Hause, Earnings profile: ability and schooling, Journal of Political Economy (May/June 1972) Part II.