An Explanation of the Educational Structure of Occupations

Theo Beckman – Ron Dekker – Andries de Gripp – Hans Heijke

Abstract. This article sets out to disclose the factors underlying the development of the educational structure of occupations. The shares of educational classes in the employment of occupations in economic sectors are taken as points of departure. The explanatory variables are technological progress and substitution processes generated by shifts in the skill structure of labour supply. For each educational class, an estimation is made with the help of a linear model, the estimation being that of Weighted Least Squares.

Introduction

This article specifies and estimates a model intended to explain the educational structure of employment in occupational classes within industrial sectors. The model is to serve as a tool for drawing up medium-term forecasts of the demand for manpower with a certain education.

The educational model fits into a larger framework, namely, the construction of a model for medium-term forecasts as part of the information system on education and the labour market being developed by the Research Centre for Education and the Labour Market (ROA). At an earlier date, Dekker, De Gripp and Heijke (1990) developed a model for making medium-term forecasts of employment ("expansion demand") for 82 occupational classes in 22 sectors of industry. The results of those forecasts are to be distributed among types of education with the help of the educational model now being developed.

The educational model is specified by defining each occupation's share in the employment of an industrial sector. The share each type of education has in employment by occupation within an industrial sector is explained in two steps. The first serves to explain the share which each level of education has in employment by occupation within an industrial sector. The second explains the shares which the various disciplines.

The authors are affiliated with the Research Centre for Education and the Labour Market (ROA), University of Limburg, Maastricht.

Received on April 9 and approved by the Editorial Board on December 10, 1991
within each level of education hold in employment by occupation and industrial sector.

Contrary to the traditional manpower-requirements approach, we do not assume fixed relationships between occupations and types of education. As in some earlier Dutch manpower studies (ANTOS 1984, NEI 1986), we assume a somewhat more flexible relationship. Another difference from traditional models is that we shall try to find explanatory variables for the changes in the employment shares of levels of education, and pool the data over occupational classes, sectors of industry and time.

The article is organised as follows: Chapter 2 takes up the factors from economic literature which are likely to influence the educational structure of occupational classes within sectors of industry. In Chapter 3, a functional form for the educational model is selected. Chapter 4 presents the data employed and the estimation results, Chapter 5 draws some conclusions and briefly evaluates the educational model.

2. The Educational Model

2.1 Classification of Types of Education and Occupation

The ROA information system of education and the labour market distinguishes only a limited number of sectors of industry, occupational classes and types of education. The classes of industry have been combined in the 22 sectors used in the sectoral forecast of the Centraal Planbureau. The occupations have been arranged in 82 occupational classes (classified by the first two digits of the occupational classification of the Central Bureau of Statistics, CBS).

This article uses the same division into industrial sectors and occupational classes as in the occupational model, thus ensuring the envisaged correspondence between the two models. For the division into occupational classes (henceforth to be called "occupations") see Dekker, De Grip and Heijke (1990).

The types of education distinguished by the first three digits of the educational classification of the CBS ("SOI") have been grouped into 59 educational categories, which we will continue to indicate as types of education. They can be distinguished by level and discipline. The levels are indicated by the first digit of the original SOI-code, running from 2 (elementary education) to 6 (scientific education). The discipline is
indicated by the second and third digits of the same code. The number of disciplines distinguished within a level of education varies from 11 (MAVO = secondary general education, lower level; LBO = junior vocational education) to 17 (higher vocational education); elementary education contains only one general category, however. Examples of disciplines are the general, technical, and medical disciplines. Among the existing types there are some which are difficult to accommodate in the 59 categories we have distinguished. We have combined these in a residual group whose development is assumed to be constant, and which may therefore be left out of the analyses.

2.2 Technological Development and Displacement

We have tried to draw up the educational model in such a way that the employment prospects of persons with a given education can be explained by technological development and the displacement processes on the labour market.

That technological advance can play an important role in the changing educational structure of an occupation within a sector is obvious. For one thing, with technological progress may come higher educational demands on new staff recruited for certain occupations or in certain industrial sectors. In that event, the functions involved, particularly those which require working with the new technology, are said to be upgraded. For another, mechanisation and automation have been directed – particularly in the last decade – towards replacing low-skilled, routine actions, thus saving on low-skilled labour. By contrast, in some occupations or industrial sectors technological progress may increase the demand for lower-educated manpower, because the higher-skilled tasks are split up and routinised. In that case, downgrading is the proper term.

So, the final effect of technological progress cannot be assessed unequivocally in advance (see also Spenner 1985). We shall explicitly include technological advance in our model, without making a priori statements about the expected development.

As criteria of technological advance in a sector we will use two regressors:
- capital intensity, measured as the ratio between the sector’s investments and value added;
- the degree of automation.

The latter is of course related to the former but, especially for the service
sectors, it might be the better indicator of technological change. Unfortunately, however, no exact data are available about the degree of automation in the sectors distinguished by us for which short-term forecasts are available. Investments by sector of industry comprise those in automation, but we have no information about the latter’s share. Therefore, we have taken the number of automation experts (occupational class 08) employed in a sector as indicator of its degree of automation or computerisation, and included it as such in the educational model as a regressor.

However, shifts in the educational structure of occupations may also be initiated by supply factors. In an easy labour market, for instance, the competition for jobs may be such that the higher qualified displace the lower skilled from their “occupational territory”. The process is called downward displacement (see Thurow 1975), and described even more aptly by Blaug as “bumping down” (De Grip 1987). The result is that people with a high level of education occupy functions previously held by those with a lower level. Typically, this type of displacement is a one-way process: only people with a higher level of education are able to push the lower-skilled from their jobs.

Economists may prefer the (neo)classical procedure, by which the phenomenon of displacement is described with the help of the relative scarcity of a certain level of education as expressed in the relative wages (see also Centraal Planbureau, 1987). However, for lack of adequate information about the wages earned by people with a given education, we will have to resort to an approximation. The assumption is warranted that a relatively generous share of a given level of education in the potential labour force will correspond to a relatively generous supply of that category of labour on the labour market. The price of such labour will be relatively low and, ceteris paribus, the people involved will be put to work more than others (in new occupations and sectors). The difference with the hypothesis of downward displacement is that, while the latter only accommodates displacement of the lower-skilled by the higher-skilled, the neoclassical approach also recognizes the possibility of the lower educated replacing the higher educated.

Despite the use of strongly aggregated data, we will try to include explicitly in the model the phenomena just described as variables explaining the shifting educational structure of occupations. To that effect we shall follow the Centraal Planbureau (1987) in adding the share of people with a given level of education in the total potential labour force as third regressor to our model. To distinguish it from the variable generated by technological developments (demand factor), we will refer to the third
regressor as "displacement variable", even though it is evidently not a displacement variable in the narrow sense, but rather an attempt at modelling substitution processes set in motion by supply factors.

3. The Econometric Specification of the Model

3.1 Introduction

As pointed out in the introduction, two steps are needed to relate employment by occupation to employment by type of education. The employment shares of levels of education will be assessed first, and next for each level the shares by discipline. The split approach has been chosen because the required level of education is probably determined by other criteria than the required discipline. We cannot know in advance, however, which has to be determined first: the required level or the required discipline. For reasons of estimation technique, we have chosen to explain the level of education as the first step: there are only six different levels against a maximum of 17 different disciplines. In this way, the initial number of shares to be broken down further is kept low. In the second step, the shares of the different disciplines within a level of education will be explained.

3.2 Employment Shares by Level of Education

As said above, the first step is to explain the shares of the various levels of education in the occupations within industrial sectors. To that end a model is used which takes the absolute volume of employment by occupation and industrial sector as given, the dependent variable being the share which a level of education takes up in the employment of an occupation within a sector. Following the approach used for the occupational model (see Dekker, De Grip and Heijke, 1990), we use a multiplicative specification:

\[ O_{kit} = \text{cons}_i \times PL_{ki}^c \times INVVA_{ki}^p \times AUT_{ij}^c \times \exp [\gamma_{kij}] \times \exp [u_{kij}] \]  

\[ i=1..82; \ k=2..6; \ j=1..22; \ t=1..4 \]  

in which:

\[ O_{kit} \] employment of occupation \( i \) in sector \( k \) in year \( t \) 

\[ PL_{ki}^c \] proportion of labour force \( i \) in sector \( k \) 

\[ INVVA_{ki}^p \] gross income \( i \) in sector \( k \) 

\[ AUT_{ij}^c \] average unit of employment \( i \) in occupation \( j \) in sector \( k \) 

\[ \gamma_{kij} \] coefficients \( i \) in sector \( k \) in occupation \( j \) 

\[ u_{kij} \] random error term
\( O_{kij} = E_{kij}/E_{ij} \) is the proportion of \( k \)-level workers (E) in the total number of people employed in occupation i within sector j.

\( PLF_k \) is the relative share of educational level \( k \) in the potential labour force.

\( INVVA_j \) is investments in sector \( j \) divided by value added in sector \( j \).

\( AUT_j \) is the proportion of automation experts in employment in sector \( j \).


\( g_{i.82} \) is occupation dummy, \( g=1 \) if the observation refers to occupation i.

\( u_{kij} \) is disturbance term (normally distributed with expectation 0).

According to equation [1], the share of educational level \( k \) in the employment of a given occupation \( i \) within sector \( j \) is determined by a constant term, the share of this educational level in the potential labour force (the displacement variable), the investments by unit of value added in sector \( j \), and the share of automation experts in this sector’s employment (the technology variables). We use dummy variables for the various occupations to take account of differences among them not expressed in the exogenous variables mentioned.

The above equation can be linearised by taking the logarithm:

\[
\ln O_{kij} = \text{const}_k + \alpha_k \ln PLF_{it} + \beta_k \ln INVVA_{it} + \kappa_k \ln AUT_j + \gamma_k g_{i.82} + u_{kij}
\]

\( i = 1..8; \quad k = 2..6; \quad j = 1..22; \quad t = 1..4 \)

An initial problem encountered is that equation [2] has to be estimated for each of the five educational levels (\( k = 2..6 \)), while the shares must add up to 1. With 6 equations for 5 shares to be estimated, the model is indeed overidentified. In principle, there are two solutions for this problem:

1. To estimate the shares for four levels, leaving the remaining unestimated. One snag of this so-called linear probability model is that no parameter values are found for the influence of the exogenous variables on the remaining level. Another is that there is no guarantee for the share of the non-estimated level to remain positive.

2. To relate the shares to a reference level, in which case it is called a distribution model. All equations contain the same regressors, but with different parameters. When such a model is estimated, the parameters of each educational level are estimated in relation to those of the reference level. The drawback of this solution is that (in this specific case) the parameter for PLF of the reference level has to be set, because PLF, unlike the other regressors, changes with the level of education. Besides, on the
An Explanation of the Educational Structure of Occupations

reference level no direct effect of the exogenous variables can be assessed with this model either. The advantage of this solution is, however, that the shares of levels remain both below 1 and positive. For the estimation, a model analogous to Parks’s multinomial logit model can be used, as was done by the Centraal Planbureau (1987). A disadvantage is that a shock in the development of the reference level occasions an inverse shock on all other levels.

A distribution model is no feasible proposition in our case, because no educational level can be found from the data that occurs in all cells of the datamatrix: only the linear probability model (equation [2]) can be estimated. On the other hand, the combination of a multiplicative model with a linear probability model is not a good choice either, because in that case a linear probability model is not consistent. The implication is that for an adequate forecast of the employment shares, the shares will have to be normalised afterwards to make them add up to 1.

The assumption is that the effect of the exogenous variables on the development of the employment shares of a certain level of education is the same for all occupations and sectors. For each occupation a dummy is indeed included to account for a constant element in the differences between the employment shares of different occupations within a sector. Moreover, the effect of exogenous variables may vary in sense and intensity from one level of education to another.

3.3 Employment Shares by Educational Discipline

A simple specification is used to determine the share of each discipline in the employment of a given occupation within a certain sector by a constant term and a trend term. No explanatory variables have been introduced because there is no complete clarity about the processes involved in the explanation of the share of the various disciplines in the employment of a given level of education within an occupation. In principle, processes could be assumed similar to those modelled in the first step, and the corresponding exogenous variables specified, but in practice it is not feasible for lack of data by discipline. The constant term has been broken down into a general, an occupation-specific, and a sector-specific constant. The specification is as follows:

\[ R_{ijkt} = \text{cons} \times \exp \left[ \gamma_1g_{ij} \times \exp \left[ \delta_1d_{ij} \right] \times \exp \left[ \tau_1i \right] \times \exp \left[ \nu_1ijkt \right] \right] \]

\[ 1 = 1.17; \quad i = 1.82; \quad j = 1.22; \quad k = 2.6; \quad t = 1.4, \]
in which:

\[ R_{ijkt} = E_{ijkt} = \text{share of people of discipline } i \text{ employed in occupation } i, \]

sector \( j \), and educational level \( k \) at time \( t \)

\[ g_{i,82} = \text{occupation dummy, } g_i = 1 \text{ if the observation refers to occupation } i \]

\[ d_{j,22} = \text{sector dummy, } d_j = 1 \text{ if the observation refers to sector } j \]

\[ t = \text{trend term (1979, 1981, 1983, 1985)} \]

\[ v_{ijkt} = \text{(normally distributed) disturbance term with expectation 0} \]

This equation can be linearised by taking the logarithm:

\[ \ln R_{ijkt} = \text{cons'} + \gamma_{ijkt} + \delta_{ijkt} + \tau_{ijkt} + v_{ijkt} \quad [4] \]

To the analogy of the first step, we estimate this equation for each discipline.

Modelling in two steps implies that each type of education is split into independent level and disciplinary aspects. In the first, the level of education is determined. Actually, a certain level sometimes implies certain disciplines, because not all levels have the same disciplines. The shifts in the educational structure in terms of disciplines are therefore additional to the changes in the disciplinary structure already implied in the model of the educational levels.

3.4 Specification of Variables

This section goes into the details of the construction of the exogenous variables.

As described in section 2.2, variable PLF (potential labour force) is a measure of the relative supply of a given level of education and thus for the wage relationships (at a given demand structure). PLF_{kt} (potential labour force with k-level of education in year t) is constructed on the basis of the bi-annual Labour Force Censuses, which break down the population between the ages of 15 and 64 by level of education.

The variable PLF represents the supply effect. Because of the substitution effect described in section 2.2 we expect this variable to have the positive sign, at any rate for the higher educational levels, and its coefficient to rise with the level of education.

We have considered the following alternative specifications of this regressor:
An Explanation of the Educational Structure of Occupations

- the share of the next higher level of education;
- the sum of the shares of all higher levels of education;
- the share of the level of education concerned divided by the sum of the shares of the higher levels;
- the share of the level itself divided by the share of the next higher level.

The first two specifications can be regarded as a more explicit modelling of the downward-displacement hypothesis. The variables represent so-called "passive substitution"; working people with a low level of skill carrying the burden of all higher levels.

The problem encountered in estimating the influence of one of these specifications is that the data vary so little. The values of the variables for the different levels of education show a continuous rising or declining pattern. As a result, the exogenous variable may behave like a trend term, which makes it hard to find a specific economic interpretation with respect to active or passive substitution. We have therefore opted for the simplest specification, namely, the share of the level of education under study.

The variable INVVA is an indication of technological advance in a given sector of industry. The variable represents, for each sector of industry, the volume of investment in equipment, transportation and engineering works over the past five years in relation to the volume of value added in the same period.

\[
INVVA_{jt} = \frac{\sum_{h=-4}^{0} INV_{jt+h}}{\sum_{h=-4}^{0} VA_{jt+h}}
\]

INV\(_{jt}\) = investments in sector \(j\) in year \(t\)
VA\(_{jt}\) = value added in sector \(j\) in year \(t\)

As stated in section 2.2, the technological developments in a sector may have consequences for the educational structure in that sector. We expect the upgrading processes to be stronger than the downgrading ones, so that for the study period the coefficient of this variable will have the positive sign for the higher, and the negative sign for the lower levels of education.

Lacking direct figures about the degree of automation in the different sectors of industry, we measure the effect of automation by the share of automation experts in the total number of people employed within a sector (variable AUT).

AUT\(_{jt}\) = Proportion of people employed in occupational class 08 (= system
analysts, statisticians, mathematicians and related specialists) in sector \( j \) in year \( t \).

We have assumed that a large proportion of automation experts in a sector of industry corresponds to a growing share of the higher educated and a declining share of the lower-skilled, because we expect automation to entail mainly upgrading processes. We therefore expect the coefficient of this variable to have the negative sign on the lower levels of education.

4. The Estimation Results

4.1 Estimation Results by Level of Education

To estimate the model, we arranged the data from the original datasets (Labour Force Censuses 1979-85) into cells. A cell is the combination of a sector of industry and an occupation; its contents represent the number of people employed in a certain occupation in a certain sector of industry. Because the data originate from a sample, they can be expected to show heteroscedasticity, which we presume to be associated with the number of working people represented in a cell. We have applied the Goldfeld-Quandt test to find out whether heteroscedasticity occurs with unweighted regression (see, among others, Judge et al. 1982). For this test the assumption is that observations can be arranged by increasing variance. From the outcomes, the zero hypothesis (no heteroscedasticity) must be rejected.

Because 150 employed people in our sample stand for approximately three persons in the original dataset, the values of very small cells are likely to be highly fortuitous. For that reason, we omit all cells of fewer than 150 persons from the analysis. In addition, we weight the remaining observations with the logarithm of the number of employed in a cell to make the observations with more people count more. The effect of the logarithmic weight becomes less as the cell increases, so that the difference in weight for large cells is not great. Moreover, we believe that weighting with the logarithm of the number of people in a cell, rather than with that number as such, provides a reasonable correction for heteroscedasticity.

Table 1 shows the estimation results. The parameters of the variable \( \ln \text{PLF} \) all have the positive sign, one cause being the separate estimation of the equations. Therefore, the parameter values need to be considered in
An Explanation of the Educational Structure of Occupations

their mutual relationship. The parameter for the lowest level appears to have the largest positive value (contrary to our expectation, but analogous to the results of the Centraal Planbureau, 1987). The high value found is probably due to the simultaneous decline of this level's share in the potential labour force (PLF) and the proportion of employed with this level of education. The parameter can be interpreted as a downward displacement effect, because a declining proportion of the lowest-skilled in the potential labour force, tantamount to a rising proportion of high-skilled people, entails a more than proportional decline of the proportion of low-skilled people in employment.

| Table 1. Estimation Results of Step 1: Employment Shares by Level of Education |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Level 2 | Level 3 | Level 4 | Level 5 | Level 6 |
| In PLF | 1.230 | 0.613 | 0.946 | 0.626 | 0.568 |
| (11.15) | (1.65) | (8.23) | (3.28) | (2.19) |
| In INVVA | -0.057 | -0.043 | 0.014 | 0.034 | 0.083 |
| (-4.12) | (-4.36) | (1.54) | (2.06) | (3.31) |
| In AUT | -0.036 | -0.012 | -0.007 | -0.021 | 0.0002 |
| (-5.50) | (-2.41) | (-1.46) | (-2.36) | (0.01) |
| F-value | 63.47 | 37.97 | 31.88 | 52.42 | 39.68 |
| R² | 0.700 | 0.536 | 0.485 | 0.712 | 0.766 |

(t-values in brackets)

Source: ROA

The parameter values of the variables In INVVA and In AUT should also be considered in their mutual relationship. In accordance with our expectations, the parameter values of In INVVA are lower for the lowest levels of education than for the higher levels. The conclusion may be, then, that employment for low-skilled workers is negatively affected by the introduction of new technology. With In AUT, the variable representing the degree of automation in a sector, we also find a stronger negative effect for the lowest than for the highest levels of education. So, with intensified automation, employment for the higher levels of education tends to increase at the expense of the lower levels.

4.2 Estimation Results by Discipline

In the second step, equation [4] is estimated for all disciplines. This model contains no explanatory variables, only a trend term and dummies for each level of education, occupation and sector of industry. When the
trend term yields no significant parameter it is set at zero. In view of the large number of regressions (59, each with some 105 estimated coefficients), we will not represent the estimation results. To judge from the R-squares of the equations, the estimation results appear to give a reasonable description of the employment structure by type of education. Some types of education appear to develop according to a strong trend (10 per cent). A fast declining trend was found with junior medical schools, lower business education, courses for driving instructors, sports coaches, etc. A fast rising trend was found with junior transport and communication schools, training for the café establishment license, and the university study fields of business administration (ir), econometrics, and actuary. These are trend-wise developments, additional to the development of the employment shares of the educational levels concerned: the two developments may either compensate or reinforce each other.

5. Conclusions

In this article we have specified an educational model in the form of employment shares by occupation within a sector of industry. These employment shares have been explained in two steps. In the first, the shares of the various levels of education in employment by occupation and sector, explanatory variables being the relative labour supply (displacement variable), capital intensity and the degree of automation (technology variables); in the second, the shares of the various disciplines within a level of education in employment by occupation and sector of industry.

The estimation outcomes of the educational model largely confirm our expectations with respect to the exogenous variables. In particular for the first step the results seem plausible enough. The technology variables are almost invariably significant, and their sign is in accordance with our expectations. The outcomes of the second step cannot be interpreted instantly, but they appear to give an adequate picture of the discipline structure of types of education, so that the model gives a reasonably accurate description of reality.

While producing plausible results, the model used so far shows some deficiencies. In particular, we need to try for a model which, despite numerous missing cells, still accounts for the relationship between the various levels and disciplines of education as well as the specific structure of the disturbance term (consisting of one component because of the sample data and another because of misspecification). For lack of a reference level for the educational model, we have not used a multinomial
An Explanation of the Educational Structure of Occupations

logit model. The specification chosen, however, has the disadvantage of not heeding the connection between the skill levels. Moreover, the introduction might be considered of a more fluctuating supply variable than the variable used in the model presented here (the potential labour force), such as the outflow from schools or the changes in it.

Notes

1 Unfortunately, for four out of the seven service sectors there in no data on investment by sector available (needed to construct NVVA). The total volume of investment in these four sectors is known, however. We therefore distributed total investment among the four sectors on the basis of total value added of these sectors. The assumption is thus that investment by unit of value added is equal for the four sectors concerned.

2 This declining trend is probably largely due to the abolition of specific training for infant-school teachers.

References


