DOCUMENT RESUME

ED 132 681 EA 008 966

AUTHOR Tinbergen, Jan
TITLE Income Distribution Over Educational Levels: A Simple Model.
PUB DATE Sep 72
NOTE 14p.; For a related document see EA 008 965

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Economic Research; *Income; *Labor Market; Labor Supply; Manpower Needs; *Mathematical Models; Multiple Regression Analysis; Occupational Information; Productivity
IDENTIFIERS *Netherlands

ABSTRACT

An econometric model is formulated that explains income per person in various compartments of the labor market defined by three main levels of education and by education required. The model enables an estimation of the effect of increased access to education on that distribution. The model is based on a production for the economy as a whole; a function from which the contribution to national product by persons in each of the five categories of manpower is considered. This production function constitutes the supply side of jobs. The demand side is derived from utility functions that contain one parameter indicating the quality of the individual, here represented by his level of education. In addition the utility function contains a variable indicating the type of job taken. A distinction is made between primary labor income and secondary income, that is, income after direct taxes. The model is based on a few more assumptions replacing lacking information. The model is tested with the aid of figures collected for the Netherlands around 1960-66. (Author/HLF)

JT/SvW/1-9-1972/28

INCOME DISTRIBUTION OVER EDUCATIONAL LEVELS: A SIMPLE MODEL

Jan Tinbergen

1. Purpose of article; main findings

In this article an attempt has been made to formulate an econometric model meant to explain income per person in various compartments of the labour market, defined by the three main levels of education and by education required. Moreover the model enables us to estimate the effect of increased access to education on that distribution. The model is based on a Cobb-Douglas-like production function for the economy as a whole; a function from which the contribution to national product by persons in each of the five categories of manpower considered. This production function constitutes the supply side of jobs. The demand side is derived from utility functions following a theory developed elsewhere [1] which contains one parameter indicating the quality of the individual, here represented by his level of education. In addition the utility function contains a variable indicating the type of job taken. A distinction is made between primary labour income and secundary income, that is, income after direct taxes. The model is based on a few more assumptions replacing lacking information. The model is tested with the aid of figures collected for the Netherlands around 1960/66. The set-up deviates, on some points to be discussed later from a more detailed model constructed also for the Netherlands [1]; in some respects it is simpler, but in one other it follows an alternative rather than a simpler method. The author believes that some aspects of the model are in need of further analysis, which he hopes to be able to carry out later. At the present stage the model seems worthy of submission to the criticism of the profession.

For what the model is worth it suggests that the ratio of primary labour incomes of the upper and the lower group considered of about 3 can be reduced to about 1,5 if instead of 3 per cent of the working force 6 per cent had third-level education and equality of incomes would require that 8 per cent of active population is able to absorb third level education.

2. Description of the model: (I) variables

For a number of variables two suffixes will be used to indicate various compartments of the labour market. The first suffix indicates the nature of the job held, by the education level best fitted to the job category. The assumption made here is that this level coincides with the upper quartile of the individual levels found in the census of population for the occupations considered. The second suffix indicates the actual level of education of the group (represented by the median of the group).

Variables showing these two suffixes are $l_{hh'}$, primary labour income; $\varphi_{hh'}$, portion of labour force working in compartment h, h'; $x_{hh'}$, labour income after direct taxes. There are also two parameters with two suffixes, π_{21} and π_{32}, indicating the ratio of productivity in job 2 with education 1 to the productivity of people with the same education but working in job 1 (for π_{21}) and the corresponding ratio for individuals on job 3 but with education 2 (for π_{32}). While for the parameters just discussed only two combinations of suffixes have been introduced (since $\pi_{11} = \pi_{22} = \pi_{33} = 1$ anyway), the number of suffix combinations is five for the variables mentioned, shown in Table 1.

Table 1. Symbols and figures of the distribution of the labour force over compartments considered, the Netherlands 1960

Level of education:		1	2	3
Job paramater:	1	φ_{11}=0.794	.	.
	2	φ_{21}=0.117	φ_{22}=0.033	.
	3	.	φ_{32}=0.026	φ_{33}=0.030

This table is not in complete agreement with reality, where small groups are found to exist where dots have been introduced in Table 1. The reason for the deviation is that it is irrational for the individuals to choose the empty compartments (cf. Section 3).

Other variables have only one suffix. This applies to F_1, F_2 and F_3 indicating the total numbers of people with education levels 1, 2 and 3, expressed as portions of the active population. Then, it applies to the exponents in the Cobb-Douglas-like production function, written as ρ_1, ρ_2 and ρ_3 (cf. again Section 3). It also applies to some coefficients where no indication of education level is at stake; thus we have two coefficients $\underline{c}_0$ and $\underline{c}_2$ in the utility function, chosen this way in order to correspond with another article on the same subject [1]. We also have three coefficients ξ_1, ξ_2 and ξ_3 to relate income after tax to primary income. Finally, we have three coefficients in what will be called the income scale offered by the organizers of production to those applying for jobs; these coefficients will be written λ_1, λ_2 and λ_3. The only variables without suffixes are total production $\underline{y}$, a constant $\underline{C}$ and multiplier $\underline{p} = 1.25$ attributed to the production factor capital. Since we count the numbers of the labour force as their portion in total population, $\underline{y}$ can also be said to represent average income and y/p average labour income.

3. Description of the model: (II) relations

The relations of the model are twenty-one. For elegance's sake we start numbering them by (0) because of the possibility this opens to present three times five equations which follow by the numbers (1) through (15). The production function (0) is written:

$$y = Cp\,(\varphi_{11} + \pi_{21}\varphi_{21})^{\rho_1}(\varphi_{22} + \pi_{32}\varphi_{32})^{\rho_2}\varphi_{33}^{\rho_3} \qquad (0)$$

The figures in brackets are the portions of the working force with education level 1 and 2, respectively, where people in the next higher job are given a weight π_{21} and π_{32} respectively. The basic assumption of the model is that people with the highest level of education are in short supply and hence all employed in level-3 jobs. The elasticities of production with regard to the numbers with education levels 1, 2 and 3, respectively are ρ_1, ρ_2 and ρ_3. As is well known, under the assumption of free competition among entrepreneurs, these exponents at the same time indicate the portions of income received by these groups. As a consequence the per capita incomes of the five categories are:

$$l_{11} = \frac{\rho_1 y}{\varphi_{11} + \pi_{21}\varphi_{21}} \tag{1}$$

$$l_{21} = \frac{\pi_{21}\rho_2 y}{\varphi_{11} + \pi_{21}\varphi_{21}} \tag{2}$$

$$l_{22} = \frac{\rho_2 y}{\varphi_{22} + \pi_{32}\varphi_{32}} \tag{3}$$

$$l_{32} = \frac{\pi_{32}\rho_2 y}{\varphi_{22} + \pi_{32}\varphi_{32}} \tag{4}$$

$$l_{33} = \frac{\rho_3 y}{\varphi_{33}} \tag{5}$$

In addition we assume that these incomes are situated at an "income scale":

$$\underline{l}_{hh'} = \underline{h}\,\lambda_1 + \underline{h}'\lambda_2 + \lambda_3 \qquad (h,\ h' = 1 \ldots 3)$$ (6) through (10)

This relation applies to the five combinations of $\underline{h}$, $\underline{h}'$ shown in Table 1; as an example we have for $\underline{h} = 3$. $\underline{h}' = 2$:

$$l_{32} = 3\,\lambda_1 + 2\,\lambda_2 + \lambda_3 \tag{9}$$

The simultaneous validity of equations (1) through (5) and (6) through (10) of course implies some relationships between some of the parameters introduced, which will be discussed in Section 5. For the moment let me state that with a high degree of accuracy the empirical material used (cf. Section 4) satisfies these conditions.

A third group of five equations indicates the relationship between income after tax $\underline{x}_{hh'}$, and primary income for each ($\underline{h}\ \underline{h}'$) considered:

$$\underline{x}_{hh'} = \xi_1 l_{hh'} - \frac{1}{2}\xi_2 l^2_{hh'} + \xi_3 \qquad (11) \text{ through } (15)$$

This relationship characterizes the tax rate system; for $\xi_1 > 0$, $\xi_2 > 0$ it reflects a progressive income tax régime.

While the preceding equations represent the supply of jobs by the organizers of production we now turn to the demand side. In principle each individual will demand for the type of job which maximizes his utility. We assume the existence of such functions in the form:

$$\omega_{hh'} = \ln \left\{ x_{hh'} + \frac{1}{2} c_o (h - 2)^2 - \frac{1}{2} c_2 (h - h')^2 \right\}$$

The assumption implies that a utility function contains:

(1) parameters characterizing the individual; in our case only $\underline{h}'$, the level of education attained;

(2) variables characterizing the possible jobs; in our case only $\underline{h}$, the level of education preferred by the organizers of production for the jobs considered and

(3) coefficients characterizing the human species, in our case only $\underline{c}_o$ and $\underline{c}_2$.

Elsewhere a number of arguments have been offered in defence of our assumptions [1], including the appearance of the term in $(\underline{h} - \underline{h}')^2$. In the same study it was found that the influence of $\underline{h}$ on utility was found to be negative for low levels of education and positive for high levels of education. In the present article this situation has been reflected by the choice of the term $\frac{1}{2} c_o (\underline{h} - 2)^2$.

Confronted by an income scale (6) - (10) and a tax regime (11) - (15) an individual with a given $\underline{h}'$ tries to maximize his utility under these restrictions. In the present model with five possible positions only we can replace this maximization process by o equations and one inequality which have to be satisfied. These have to express the situation for each of the three types of people. Those with primary education will tend to prefer a job $\underline{h} = 1$, depending, however, on the net income attached to such a job; since there is a surplus of persons with primary education only, part of the group will take ta higher job, $\underline{h} = 2$. In order that these two situations coexist, it must be indifferent for those with $\underline{h}' = 1$ whether they take $\underline{h} = 1$ or 2; that is

$$\omega_{11} = \omega_{21}$$

Upon substituting into this equality $\underline{x}$ by $\underline{l}$ and $\underline{l}$ by its expression in terms of the income scale we obtain:

$$x_{11} + \frac{1}{2} c_o = x_{21} - \frac{1}{2} c_2 \tag{16}$$

or:

$$\frac{1}{2} c_o + \frac{1}{2} c_2 = \xi_1\lambda_1 - \frac{3}{2}\xi_2\lambda_1^2 - \xi_2\lambda_1\lambda_2 - \xi_2\lambda_1\lambda_3 \tag{16'}$$

Similarly for those who have secondary education ($\underline{h}' = 2$) we must have

$$\omega_{22} = \omega_{32}$$

or

$$x_{22} = x_{32} + \frac{1}{2} c_o - \frac{1}{2} c_2 \tag{17}$$

identical with:

$$-\frac{1}{2} c_o + \frac{1}{2} c_2 = \xi_1\lambda_1 - \frac{5}{2}\xi_2\lambda_1^2 - 2\xi_2\lambda_1\lambda_2 - \xi_2\lambda_1\lambda_3 \tag{17'}$$

For people with $\underline{h}' = 3$ there will be sufficient jobs $\underline{h} = 3$, and the wage and tax scales will be such as to make this preferable to them in comparison with $\underline{h} = 2$:

$$\omega_{33} \geq \omega_{23} \text{ or } x_{33} + \frac{1}{2} c_o \geq x_{23} + \frac{1}{2} c_o - \frac{1}{2} c_2$$

or:

$$\frac{1}{2} c_o + \frac{1}{2} c_2 \geq -\xi_1\lambda_1 + \frac{5}{2}\xi_2\lambda_1^2 + 3\xi_2\lambda_1\lambda_2 + \xi_2\lambda_1\lambda_3 \tag{17''}$$

For the numbers of people in each of the compartments in Table 1 we must have the balance equations

$$\varphi_{11} + \varphi_{21} = F_1 \tag{18}$$

$$\varphi_{22} + \varphi_{32} = F_2 \tag{19}$$

$$\varphi_{33} = F_3 \tag{20}$$

The 21 equations (0) through (20) enable us to solve for the unknowns $\underline{y}$, $\underline{l}$ (5 in number), $\underline{x}$, φ (also each 5 in number) and the income and tax scales' parameters (λ and ξ, where ξ_3 will appear to be irrelevant and hence 5 unknowns only). We consider as given, for the time being, the 5 coefficients ρ_1, ρ_2, ρ_3, π_{21} and π_{32} of the production function, the income portion of capital p (determined by past savings), the coefficients $\underline{c}_0$ and $\underline{c}_2$ of the utility function and the number of people (expressed as portions of the labour force) $\underline{F}_1$, $\underline{F}_2$ and $\underline{F}_3$.

The solution of this "analytical" or "explanatory" problem can be found along the following path. The five $\underline{l}$ can be eliminated between (1) - (5) and (6) - (10). The "mixed" φ_{21} and φ_{32} can be expressed in terms of the three $\underline{F}$ and the "homoneneous" $\varphi_{11}, \varphi_{22}$, whereas φ_{33} is already solved with the aid of equations (18) - (20). Both λ_1 and λ_2 can be expressed in terms of production function coefficients and homogeneous $\varphi_{11}, \varphi_{22}$ by subtracting (6) from (7), (8) from (9), (7) from (8) and (9) from (10):

$$\lambda_1 = \frac{(\pi_{21} - 1)\ \rho_1 y}{\varphi_{11} + \pi_{21}\ (F_1 - \varphi_{11})} = \frac{(\pi_{32} - 1)\ \rho_2 y}{\varphi_{22} + \pi_{32}\ (F_2 - \varphi_{22})} \tag{21}$$

$$\lambda_2 = \frac{\rho_2 y}{\varphi_{22} + \pi_{32}(F_2 - \varphi_{22})} - \frac{\pi_{21}\rho_1 y}{\varphi_{11} + \pi_{21}\ (F_1 - \varphi_{11})} = \frac{\rho_3 y}{F_3} - \frac{\pi_{32}\rho_2 y}{\varphi_{22} + \pi_{32}\ (F_2 - \varphi_{22})} \tag{22}$$

Equations (21) and (22) do not contain $\underline{y}$ anymore and hence represent two equations in φ_{11} and φ_{22}.

4. Estimation of the non-observable entities

In Sections 2 and 3 the model has been presented as an analytical or explanatory model, meaning the usual presentation of models, where the unknowns are the target variables and the irrelevant variables in the terminology used previously [2]. The same model can be used as a policy or planning model, where the unknowns are the instrument and the irrelevant variables. Some of the entities involved can still be interpreted differently, and interpretations are a question of taste in some cases. Thus, the λ can be conceived as self-adjusting market characteristics or as consciously government-used instruments.

Independently from these considerations the problems of measurement may be seen as a third category of using the model. I will call this the statistico-econometric use of the model. Direct statistical observation supplied us with the values of the $\underline{l}$, the $\underline{x}$, the φ and $\underline{F}$, and hence of the ρ and the π. The ξ, but also the λ_i were estimated with the aid of (extremely simple) multiple regression techniques, however. Thus we found:

$$l = 1.54\ h + 5.0\ h' - 0.9 \tag{23}$$

and

$$x = 1.5\ l - 0.030\ l^2 - 3.4 \tag{24}$$

meaning that

$$\lambda_1 = 1.54, \quad \lambda_2 = 5.0 \text{ and } \lambda_3 = -0.9$$
$$\xi_1 = 1.5, \quad \xi_2 = -0.030 \text{ and } \xi_3 = -3.4$$

Since the number of unknown coefficients to be estimated was equal, in both cases, to the number of observations, the reason why these regression equations did nevertheless not show multiple regression coefficients exactly equal to one is that the "scales" are somewhat of a straightjacket to reality, but not much of it. The fits are quite good.

Finally, it was possible to reveal, with the aid of equations (16') and (17'), the utility function used and to find that, granted our theory of the utility function, the best fits yield

$$c_0 = 0.63 \quad ; \quad c_2 = 2.84 \qquad (25)\ (26)$$

These coefficients also appeared to fulfil inequality (17'').

The importance of having these estimates is that they in principle enable us to determine the social optimum (requiring one more assumption on how to determine social welfare,given individual welfare (utility) functions).All this has been discussed in the earlier study referred to already several times. We will discuss a few similar rather ambitious applications of our model in Section 5. A survey of all figures found and used will be found in Table 2.

le 2. Values of observed as well as non-observed but estimated variables and parameters in the model (The Netherlands, 1960/6)

mbol	Values of h, h'	Units	11	21	22	32	33	Average
h'	Primary labour income	hfl 1000	6.5		9.7		19.0	7.23
h'	Income after tax	"	5.7		8.3		14.0	
								Total
h'	Frequency	per cent	79.4	11.7	3.3	2.6	3.0	100
'	Frequency of educ.levels	" "	91.1		5.9		3.0	100
h'	Productivity ratio		1	1.27	1	1.13	1	
'	Exponent Cobb-Douglas production function		0.648		0.088		0.064	0.80

Primary income scale $l_{hh'} = 1.54\,\underline{h} + 5.0\,\underline{h}' - 0.9$

Income after tax scale $x_{hh'} = 1.5\,l_{hh'} - 0.030\,l_{hh'}^2 - 3.4$

Coefficients of utility function $c_o = 0.63 \qquad c_2 =$ [illegible]34

Utility function $\omega = \ln\left\{x_{hh'} + \frac{1}{2}c_o(h-2)^2 - \frac{1}{2}c_2(h-h')^2\right\}$

e on sources: Figures for total primary income of various social groups: Nota over de inkomensverdeling, Bijlage 15 Miljoenennota 1970. (Memorandum on income distribution, Annex 15 to Budget Proposals 1970), The Hague, Parliamentary documents, State Publ. Office, 1969.
Estimate of labour income contained in total primary income: Inkomensverdeling 1966 en vermogensverdeling 1967. (Income Distribution 1966 and Distribution of Wealth 1967), Central Bureau of Statistics, The Hague, 1971.
Income after tax derived from "Nota". Frequency distribution of working force over h and h': Census of Population, 1960, Part 8, Central Bureau of Statistics, The Hague, 1969.

5. Some uses of the model and supplementary relations needed

The essence of our model, as we see it, is that it shows the influence of education on income distribution (primary as well as after tax). The observed fact that l_{33} is about three times l_{11} and l_{22} about twice l_{11}, is linked, in our model, with the relative scarcities of people with different levels of education. Relative scarcities can be expressed by the ratios $\frac{\rho_1}{F_1}$, $\frac{\rho_2}{F_2}$ and $\frac{\rho_3}{F_3}$. In the observed situation these ratios are 0.71, 1.50 and 2.13, whose values show relative sizes corresponding fairly closely with the relative values of l_{11}, l_{22} and l_{33}, being 6.5, 9.7 and 19.0. It is also easily found that the values $\varphi_{21} = \varphi_{32} = 0$ together with the assumption

$$\frac{\rho_1}{F_1} = \frac{\rho_2}{F_2} = \frac{\rho_3}{F_3} \tag{27}$$

are compatible with our model and then yield $l_{11} = l_{22} = l_{33}$ with $\lambda_1 = \lambda_2 = 0$.

Complete equality of incomes is possible provided that the numbers F_1, F_2 and F_3 can satisfy equations (27). Since $F_1 + F_2 + F_3 = 1$, the values of F corresponding with (27) are:

$$F_1' = 0.81 \qquad F_2' = 0.11 \qquad F_3' = 0.08 \tag{28}$$

Whether or not such equalization of incomes is possible then becomes a question of whether indeed 8 % instead of 3 % of the working force command the capabilities to accomplish a third-level education.

Assuming that such capabilities are not available with 8 % of the population, but, say with at a maximum 6 %, the problem may be formulated how changes in the F affect income distribution according to our model. In an attempt to solve this problem we meet a number of difficulties. To begin with, we cannot consider the π_{21} and π_{32} to remain constant with changes in relative productivity of the groups φ_{11}, φ_{22} and φ_{33}. Rather must we expect these π to become 1 in case the productivities just mentioned become equal as a consequence of equal scarcities of the three types of manpower considered. From a number of attempts made it seems that the path along which the π change with changing scarcities is relevant, at least if we want to maintain all other features of our model, especially the coincidence of the income scales given by equations (1) - (5) with those given by (6) - (10). Clearly it is conceivable that this coincidence does not persist.

Indicating the variables of our new problem with two primes, and asking what income distribution will result from a given rise in F_3'' while maintaining the coincidence of the two income scales just discussed, we were able to construct an extended model and solve it for $F_3'' = 0.06$. The π were replaced by the relations:

$$\pi_{21}'' = \left(\frac{\rho_2}{F_2''} \frac{F_1''}{\rho_1} \right)^{\sigma} \tag{29}$$

$$\pi_{32}'' = \left(\frac{\rho_3}{F_3''} \frac{F_2''}{\rho_2} \right)^{\tau} \tag{30}$$

where σ and τ were chosen so as to give the observed values $\pi_{21} = 1.27$ and $\pi_{32} = 1.13$ for $F''_{\underline{h}} = F_{\underline{h}}$. It appeared that $\sigma = 0.32$ and $\tau = 0.34$ fulfil these conditions. Further experiments with various sets F_1'', F_2'' yielded one possible solution for $F_1'' = 0.845$ and $F_2'' = 0.095$ with the $\varphi_{\underline{hh}'}$, as given in Table 3.

Table 3. Values for $\varphi''_{\underline{hh}'}$

h ↓ \ h' →	1	2	3
1	0.609	.	.
2	0.246	0.077	.
3	.	0.018	0.060
Total	0.845	0.095	0.060

In addition it was found that in this solution

$$\frac{1_{33}''}{1_{11}''} = 1.44$$

meaning a reduction to one half of that ratio. Inequality has been reduced even more, of course, since the frequencies of Table 3 have become less unequal simultaneously. The impression gained from the trial-and-error process carried out is one of considerable sensitivity of the solution to small changes in the values for F_1'' and F_2'' and the need for further theoretical exploration of the model presented. The model would appear anyway to be an improvement of a previous model given elsewhere, which does not permit the mixed $\varphi_{\underline{h}\,\underline{h}'}$ (i.e. $\underline{h} \neq \underline{h}'$) to become zero and cannot at all be valid in a considerable range around this zero value.

Another question arising is whether the tax rates have to be changed in order that conditions (16), (17) and (17'') remain valid. We will not go into this question here, however, but rather leave it to later occasions. It appears that the solution of this latter problem requires an income tax structure different from (11) - (15) which is sort of straitjacket (just as equations (6) - (10))which need not always be assumed.

References

[1] J. Tinbergen, "An Interdisciplinary Approach to the Measurement of Utility or Welfare", R.C. Geary lecture 1972, to be published.

[2] -----, Economic Policy: Principles and Design, Amsterdam 1956.