3

Some examples of 'price equations'

3.1. Theory and terminology used

In this book incomes are considered to be *prices* paid for various types of production factors, mainly labour of different qualifications. Capital incomes, having been discussed by many authors, are only touched upon occasionally (cf. Chapter 9). Our emphasis will be on the *compartments of the labour market* which in principle can be characterized by any number of aspects and by the *degrees* to which these aspects have to be present. Often the aspect of *years of schooling* will be used and a distinction will then be made between schooling *required* for the jobs in the compartment and *available* schooling. This implies that already this single aspect gives rise to two indicators or indexes needed to identify the compartment.

Let us first consider one single compartment and the price prevailing in it. Applying customary economic theory we will try to explain that price by the interaction between demand and supply. This implies that two relationships are being introduced, known as the *supply equation* and the *demand equation.* Both contain the price and the quantities 'bought' and 'sold'; the latter two are supposed to be identical and will be called the 'employment' of the labour type considered. As a rule the relations also contain other variables. Those occurring in the supply equation are usually called *supply factors*, and those occurring in the demand equation, *demand factors.* In a model describing one single market, price and employment will be considered to be the endogenous variables and the supply and demand factors to be the exogenous variables of the model. Solving for the two endogenous variables we obtain what we will call the *price equation* and the *employment equation.* The former expresses the price and the latter expresses employment in terms of supply and demand factors. These two equations may also be called the *reduced forms* of the market equations.

So far our considerations about a single compartment. However, our interest is in income *distribution,* in the distribution of prices and employment volumes among the various compartments. For a small number of compartments income distribution may be described by a list of all price and employment equations. A clearer picture of distribution will be obtained, however, by choosing a smaller number of *inequality measures,* especially when a large number of compartments is being studied. A large number of inequality measures are currently being used. The simplest among them are one or more ratios or differences between prices. More complicated ones are such concepts as the standard deviation of the price distribution, where employment figures are the frequencies, or the Lorenz inequality coefficient shown in Chapter 2. In a general way we may state that inequality measures are obtained by applying some operator to prices and employment figures; and with the aid of price and employment equations, the result will be expressed in terms of corresponding functions of supply and demand factors. In the simplest cases these functions will be measures of the level and distribution of supply and demand factors.

In the subsequent sections three examples will be given of this approach, tailored to the data made available by Lydall [42], Chiswick [16] and Soltow [60], all of them supplemented with additional data from other sources.

3.2. A cross-section test with national figures

The test to be discussed here is applied to income distribution figures published by Lydall [42] and referring to the ratio of the fifth percentile (from the top) to the median of Lydall's standard income distribution, where only earnings have been considered. This implies an underestimation of income inequality, although less so than many think: in developed countries three-quarters of primary income inequality (and much more of income-after-tax inequality) is due to labour income inequality.

The measure chosen for the only supply factor considered consists of *enrolment* in primary and secondary schools as a percentage of the age group from 5–19 years. A high enrolment figure is supposed to affect inequality negatively since it raises competition for qualified and lowers competition for unqualified labour. Alternatively, inequality in schooling, expressed by its Lorenz coefficient in a 1971 publication by Chiswick [15], has been introduced. As a second alternative, higher education enrolment per 1000 aged 20–29 has been used.

Demand distribution has been represented alternatively by the percentage of the economically active population in manufacturing, by the adult male labour in agriculture as a percentage of total male labour and by GNP per capita. The philosophy behind these alternative dummy variables is that more industrialized, or less agrarian, or simply richer countries will demand relatively more qualified labour. In the next section this idea will be pursued further and refined.

In this chapter differences between the technological level of countries or American states have been neglected. This subject will be taken up in Chapters 5 and 6.

One further experiment suggested in Chiswick's study [15] has been undertaken, namely the addition as an explanatory variable of the rate of growth of GNP per capita.

Table 3.1 shows some of the results obtained.

Before discussing some of my results I want to point out that, unfortunately, many of the regression coefficients found are highly *unstable*, depending on the choice of further explanatory variables or due, sometimes, to multicollinearity among the explanatory variables. In some cases this can be avoided by introducing *a priori* values or ratios of values for the regression

Table 3.I

Some regression equations obtained from cross-section tests explaining income inequality x; R = correlation coefficient, corrected for degrees of freedom.

No.	Regression coefficient for (· not included)								$\bar{R}$	Theory[a]	No. of countries included
	y	y'	z	u	v	w	w'	$\dot{w}'$			
1	.	0.34	.	.	.	.	−0.016	87.3	0.75	Chiswick	9
2	−4.4	.	.	.	.	.	.	.	0.72	Lydall	15
3	−3.6	.	−1.1	.	.		.	.	0.64	Lydall	15
4	−1.74	.	.	−5.85	.	−0.019	.	.	0.82	Tinbergen	15
5	−2.24	.	.	−6.23	.	.	.	.	0.78	Tinbergen	15
6	−3.36	.	.	.	.	−0.031	.	.	0.74	Lydall	15
7	−0.89	.	−1.3	−6.31	.	.	.	.	0.76	Tinbergen	15
8	−0.42	.	.	.	+1.72	.	.	.	0.54	Lydall	15
9	−4.6[b]	.	.	+2.3[b]	.	.	.	.	0.58	Tinbergen	17

[a] Cf. text.

[b] Ratio of coefficients, chosen on a priori grounds, 2 : −1.

Symbol definition :		*Source :*
x :	Fifth percentile, as per cent of median of 'standard income distribution' (Lydall).	Lydall [42]
y :	Enrolment (combined primary and secondary) as per cent of age group 5–19 years (indicator 23).	UNRISD [77]
y' :	Lorenz coefficient of schooling, in per mille.	Chiswick [15]
z :	Higher education enrolment per 1000 aged 20–29 (indicator 27).	UNRISD [77]
u :	Per cent economically active population in manufacturing (ISIC divisions 2,3) (indicator 55).	UNRISD [77]
v :	Adult male labour in agriculture as per cent of total male labour (ISIC division 0) (indicator 50).	Lydall [42] UNRISD [77]
w :	GNP per capita (1959/61), in 1960 US $, at parity rate (indicator 69).	UNRISD [77]
w' :	Average per capita GNP in 1967 prices for 1950/60.	Chiswick [15]
$\dot{w}'$:	Percentage change in GNP per capita in constant 1967 prices (dollar equivalent) from 1950 to 1960.	Chiswick [15]

coefficients for some variables. For the testing of my own theory this is possible, for instance, if demand for and supply of some type of qualified labour are expressed in the same units, for instance, percentages of active population.

From Table 3.I we may conclude that all cases shown are unanimous about the algebraic sign of the level of education: it is negative, meaning that more education will reduce income inequality. The order of magnitude does not vary too much either, except in equation 8, where the inclusion of v (percentage in agriculture), inspired by one of Lydall's suggestions, takes over the role of y but yields a poor correlation. The influence of u (percentage in manufacturing) appears remarkably stable, except where a positive sign has been imposed (following my own theory), but with hardly more success than Lydall's case. Where two variables of the level of education (y and z) have been introduced, the relative influence of each is open to doubt but the joined influence less so. Wherever the variable u has been introduced freely, it obtains the wrong sign from my point of view and takes over part of the influence of education.

In order to answer the question of what increase in education will be needed to reduce income inequality in the socially more advanced countries to one-half of its present level, we have to insert $\Delta x = -50$. The answer will depend on the relative change in higher (z) as compared to 'lower' (y) education. Assuming that only a portion of those who receive more 'lower' education – we assume 6 per cent – will be able to follow higher education, we have $\Delta z = 0.6\Delta y$. The results for Δy and Δz (when considered) are found in Table 3.II.

In order to appraise the feasibility of the changes found, we must compare these changes to the actual figures already attained. The highest figures for primary and secondary education enrolment given by UNRISD are 81 for the USA, Canada and Belgium; Sweden and Denmark show 69 and 71, respectively, Germany (F.R.) 72, the Netherlands 75, France 76. Some of the changes shown in Table 3.II are impossible, therefore, while others seem to be *within reach*. Enrolment for higher education is, of course, quite a bit lower.

Table 3.II
Increase in education enrolment needed to reduce income inequality by one-half.

Equation	Δy (%)	Δz (‰)
2	11	.
3	12	7
4	29	.
5	22	.
6	15	.
7	30	18
8	33	20
9	11	.

3.3. Two time series tests

Another test was based on time series data available for the Netherlands and Norway; the latter country is represented by a sample studied by Soltow [60]. Indicators available for income distribution and supply and demand factors have been given the same main symbols as in the preceding set of computations, but with a varying number of primes. Table 3.III shows the results.

Again we may ask what change in education is needed in order to reduce to one-half the income inequality of 1960. For the Netherlands this means $\Delta x'' = -0.3$. Assuming that in the future almost 90% of the population will have secondary education ($\Delta z' = 500$), we find that $\Delta z'' = 54$, which means that enrolment in higher education should be *doubled* in comparison to the present situation. For Norway we find that $\Delta x''' = -16$ requires also *doubling* z''' (from 32.10^{-5} in 1960 to 64.10^{-5}). In Norway so far z''' has doubled every thirty years. In considering these provisional results we should remember that their margin of error is very large.

Apart from education some other of the explanatory variables may also be partly or indirectly considered as action parameters. Thus, Chiswick's equation 1 (Table 3.I) suggests that a country's *rate of growth* exerts an influence on income distribution. With a lower rate of growth income distribution

Table 3.III
Time series tests based on a demand–supply theory; corrected correlation coefficient.

Country	Period	Regression equation	$\bar{R}$
Netherlands	1920–1960	$x'' = 0.6 + 0.00199u' - 0.00043z' - 0.00156z''$	0.78
Norway	1875–1950	$x''' = 43.8 - 20.000\dfrac{z'''}{u''} + 0.016\,p$	0.90

Definition of symbols:		*Source:*
Netherlands		
x''	Relative average absolute deviation (of individual incomes from average income)	CBS [11]
z'	Enrolment in secondary and scientific education per 1000 of age groups concerned	CBS [13]
z''	Enrolment in scientific education per 1000 of age group	CBS [13]
u'	Active population in manufacturing (weight 1) and services (weight 3) per 1000	CBS [13]
Norway		
x'''	Gini coefficient of income inequality, Østfold and Vestagter (for 1960 Østfold only)	Soltow [60]
z'''	Number of students passing final university exam divided by total population	Norges HS [47]
u''	Percentage of active population outside primary industries	Norges HS [47]
p	Ratio of wealth to income	Soltow [60]

will be less unequal. A reduction by 50 points in x, meaning, as I said, halving the inequality of the socially advanced countries, will be obtained if the *rate of growth of income* per capita is reduced by 0.6 per cent. The other measures used in Chiswick's article [15] for income inequality [Lydall's P(10) and P(75)], when halved, would require, according to the corresponding equations, reductions in growth rates of 1.4 and 0.8 per cent, that is, figures of the same order. The reader should again be reminded of the sensitivity of the coefficients to the inclusion of other explanatory variables.

The same equation suggests that a reduction of inequality to one-half of its existing value could be obtained by an *increase in per capita GNP* of $ 3000, or by more than doubling the 1960 American GNP per capita. I already pointed out that a difference of opinion exists among the authors discussed, even about the direction of a change in income 'needed'; this can be demonstrated with the aid of some material to which I will return in the next section.

3.4. Cross-section studies for parts of countries*

3.4.1. Introduction

Three studies by Americans, namely T. Paul Schultz [58], Leland S. Burns and H. E. Frech III [9], and Barry R. Chiswick [16] which are based on an interesting and large amount of information, are the basis for the present section. Some material selected and processed by the present author is also included.

The data pertains to subdivisions of three countries: the United States, Canada, and the Netherlands. Although the authors mentioned adhere to theories of income distribution somewhat different from my demand–supply theory [63], their material can also be used to test the latter, subject to some assumptions. The material added by my own modest extension seems to fit the purpose somewhat better, however. One of the points of focus of this section therefore consists of attempts to give practical shape to the introduction of variables supposed to represent demand. These attempts will be followed by still others in Section 3.5.

One condition to be fulfilled for any attempt to test the demand–supply theory is that the geographical units compared in a cross-section or time series analysis be *large enough* to contain both the demand and the supply location. For commuters there is a distinction between the place where they work (and where the demand is exerted) and the place where

* I want to express my sincere thanks to my collaborators A. ten Kate and H. Visscher for programming many of the calculations used in this section.

the supply is shown). This implies that cross-section studies using single municipalities, such as the Burns–Frech study and some of T. P. Schultz's investigations, may lead to unreliable results. For that reason I have preferred to use data for only the (eleven) provinces of the Netherlands, as was also done by Schultz.

As already mentioned, this section deals with cross-section analyses for three countries. The figures refer to the states of the United States (Chiswick), the provinces of Canada (same author) and a number of municipalities (Burns and Frech), the socio-geographic areas and the provinces of the Netherlands (Schultz, Tinbergen). Burns and Frech, in particular, chose the 71 largest municipalities, Schultz 88 selected at random, and both Schultz and I took the eleven provinces of my country. The advantage of the type of material chosen consists of homogeneity in cultural and other respects, partly unknown even, which does not exist in cross-section studies among widely differing countries as carried out by Lydall [42] and in Section 3.2. This homogeneity is also lacking in time series studies because of changes both in the system of education and in the technology of production.

There are also disadvantages connected with cross-section studies within a single country; one has been mentioned already: commuters do not always work and live in the same geographical unit. Another is that variations of variables within one country, especially a small country, may be so restricted as to be a hindrance to extrapolations, which are the main instruments for arriving at the more interesting answers we want to derive from our studies.

The variables used in this subsection have been listed in Table 3.IV. Except for the last column the demand index used was the average income of the geographical units considered. Schultz did not use this variable. In Subsection 3.4.5 the better demand index Y'' will be used for the Netherlands and similar data for the United States will be used in Section 3.5.

Table 3.IV
List of variables used by the authors quoted (USA = United States of America; CDN = Canada; NL = Netherlands).

Symbol	USA and CDN	NL		
	Chiswick	Schultz	Burns and Frech	Tinbergen
X	X: Variance of nat. logs of income in \$ 1000	X': Concentr. ratio of income	X': Concentr. ratio of income	X'': Highest decile of income
Y	Y: Average of nat. logs of income in \$ 1000	*	Y': Income in Dfl. 1000	Y'': Demand index[c]
Z	Z: Average number of years of schooling for males over 25	Z': Males 40–64: % with higher education[a]	Z'': Years of schooling[b]	Z''': Per cent of active population with secondary and higher education
U	U: Variance in number of years of schooling for males over 25	*	U': Concentr. ratio of schooling	U'': Per cent of active population with higher education
V	V: Nat. log of Y_0 (income at zero schooling)	*	*	*

Note: Capital letters are used for variables in units indicated; lower case letters will be used for 'normalized' variables (i.e., average = 0, standard deviation = 1). Asterisk means: variable not used.

[a] For 1960: percentage of active males with higher education.

[b] Total population.

[c] Defined in Section 3.4.5.

3.4.2. Using Chiswick's material for the United States

For each of the data collections analysed we used two ways of measuring the variables used: the 'natural units' as indicated in Table 3.IV, and the normalized units (with zero average and unit standard deviation), the latter being indicated by lower case letters. We attempted to study the structure of relationships by comparing regression coefficients for the same variable in different combinations with other variables. Chiswick's material on the USA was used for Table 3.V.

Table 3.V
Regression and multiple correlation coefficients R found for different combinations of variables explaining income inequality x.

No.	Regression coefficients for				R	
	y	z	u	v		
1	−0.79	.	.	.	0.79	1 expl. var.
2	.	−0.73	.	.	0.73	1 expl. var.
3	.	.	+0.48	.	0.48	1 expl. var.
4	.	.	.	−0.86	0.86	1 expl. var.
5	−0.60	−0.23	.	.	0.80	2 expl. var.
6	−0.71	.	+0.25	.	0.825	2 expl. var.
7	+0.08	.	.	−0.94	0.86	2 expl. var.
8	−0.82	+0.15	+0.31	.	0.83	3 expl. var.
9	+1.25	−0.67	.	−1.58	0.93	3 expl. var.
10	+0.65	.	+0.42	−1.38	0.94	3 expl. var.
11	+1.02	−0.33	+0.315	−1.52	0.94	4 expl. var.

Source: [16] Table 3-3.

We did not use all the variables shown in Chiswick's study, for instance not his variable $\hat{r}$, the rate of return on education derived per state from the regression, in that state, of income on schooling. My feeling is that its use duplicates the variables Z and U, since Chiswick's (and Mincer's) theory is that everybody's choice of length of schooling is partly based on $\hat{r}$. It seems that indeed $\hat{r}$ is superfluous, even statistically; there appears to be complete multicollinearity in the set $(x, y, z, u, v, \hat{r})$.

The following conclusions seem warranted:
The influence exerted by variables u (education inequality) and v (representing other influences on income, such as innate capabilities) are stable; variable v always raises considerably the correlation coefficient. The contribution of u is less important, but stable. The influence of y, taken here to represent the demand for qualified manpower, looks uncertain since positive as well as negative regression coefficients are found. Negative coefficients occur when and only when v is excluded. The cases with the highest multiple correlation coefficients show a positive regression coefficient for y. The influence exerted by variable z is negative in most cases. These statements induce me to select case no. 11 as the most satisfactory relationship found with the aid of Chiswick's material.

To obtain natural units we must divide the corresponding symbols by their standard deviations, given below (Source: [16] Table G-5): $\sigma_x = 0.12$; $\sigma_y = 0.23$; $\sigma_z = 0.79$; $\sigma_u = 3.17$; $\sigma_v = 0.29$; the relation then becomes

$$X = 0.532Y - 0.050Z + 0.012U - 0.629V. \qquad (3.1)$$

As an *illustration* of the influence which a higher level and a more equal distribution of education may exert, we assume an increase in years of schooling of 2 and a reduction of its variance of 4; such changes would lead to $\Delta X = -0.100 - 0.048 = -0.148$. Since the average value of X, that is $\bar{X} = 0.79$, this represents a very modest reduction of inequality in income in the United States; it *reduces the standard deviation of incomes* from $\sqrt{0.790}$ to $\sqrt{0.642}$ or from 0.89 to 0.80 or *by only 10 per cent*. As we shall see in the case of the Netherlands, the coefficients for Z and U may become larger, however, if Y is replaced by a better measure for demand.

3.4.3. Using Chiswick's material for Canada

Chiswick has collected for Canada the same material as for the USA. Some of the results obtained with its aid are given in Table 3.VI. Here we see that the influence exerted by y and

Table 3.VI
Regression and multiple correlation coefficients R found for different combinations of variables explaining x.

No.	Regression coefficients for				R	
	y	z	u	v		
1	−0.62	.	.	.	0.62	1 expl. var.
2	.	−0.54	.	.	0.54	
3	.	.	−0.15	.	0.15	
4	.	.	.	−0.67	0.67	
5	−0.55	−0.09	.	.	0.625	2 expl. var.
6	−0.85	.	+0.38	.	0.68	
7	+0.08	.	.	−0.74	0.67	
8	−1.93	+0.90	+0.91	.	0.76	3 expl. var.
9	+1.17	−0.49	.	−1.48	0.72	
10	+0.59	.	+0.82	−1.83	0.86	
11	+0.10	+0.27	+0.92	−1.61	0.86	4 expl. var.

Source : [16] Table 3-12.

z is unstable, whereas that exerted by u and v is relatively stable. Also, inclusion of u or v considerably raises the correlation coefficient. Transforming equation 11 into one with the units used by Chiswick and mentioned in Table 3.I, we obtain

$$\frac{X}{0.09} = 0.10\frac{Y}{0.21} + 0.27\frac{Z}{0.78} + 0.92\frac{U}{1.08} - 1.61\frac{V}{0.26},$$

or

$$X = 0.043Y + 0.031Z + 0.077U - 0.56V. \qquad (3.2)$$

In contrast to the result for the United States, there is a positive influence of the average level Z of education on income inequality X; this implies that the average level may already be too high. A possible explanation may be in the fact that in Canada education is obligatory to a larger extent than in the United States; at least for Great Britain this argument is used by Chiswick [14] and in this respect Canada is probably somewhat closer to Britain than the United States.

Considering that $\bar{U} = 10.69$, we may think of a reduction in the inequality of schooling as a means to reduce income inequality and estimate the influence of $\Delta U = -5$. This would mean that the standard deviation in years of schooling is reduced from $\sqrt{10.69}$ to $\sqrt{5.69}$ or from 3.27 years to 2.39 years. We obtain

$$\Delta X = -0.385. \tag{3.3}$$

Since $\bar{X} = 0.63$, this brings inequality as measured by X to less than one-half of its present value; but when measured as a standard deviation in the natural logarithms of income it falls from $\sqrt{0.63}$ to $\sqrt{0.245}$ or from 0.794 to 0.495, a *reduction by 38 per cent only.*

A common feature of the equation found for both the United States and Canada is that raising the Y_0, which stands for the factors other than schooling which determine an individual's productivity, reduces inequality in about the same way. This may in part reflect the influence of the 'environment', including the influence of the education of the parents. If this interpretation is correct, the long-run influence of education may be considerably stronger than the direct influence estimated.

3.4.4. Research on the Netherlands by T.P. Schultz and by L.S. Burns – H.E. Frech III

Schultz's contributions [58] to the explanation of income inequality (p. 352) consist of having assembled a vast collection of statistical data, for 11 provinces, for 75 regions, and for 88 municipalities selected in a random sample (pp. 339/340) and of having analysed various relations in order to explain changes over time with the aid of various explanatory variables. He has also studied cross-section data. For this publication the latter are the more relevant analyses. Income inequality among regions as well as among provinces, as measured by their concentration ratios, has been explained by a variety of variables, including the level of education, for which Schultz found a positive influence. No use is made of demand factors, which prevents us from testing the demand–supply theory. The other explanatory variables include the number of taxpayers, unemployment and wealth. The best results are obtained for the most recent year studied by him, 1958, and for the provinces. This seems to confirm the viewpoint that the geographical units should not be chosen too small. With the aid of the education level (measured as the percentage of active population having had higher education) a corrected correlation coefficient of 0.89 is obtained. This result comes close to my own results, to be discussed in Section 3.4.5.

Burns and Frech used the figures for 71 of the larger municipalities. Their material enabled me to compute Table 3.VII, where the symbols are those explained in Table 3.IV.

Table 3.VII
Regression and multiple correlation coefficients R found for different combinations of variables explaining income inequality x'.

No.	Regression coefficients for			R	
	y'	z''	u'		
1	−0.91	.	.	0.91	1 expl. var.
2	.	−0.50	.	0.50	
3	.	.	−0.68	0.68	
4	−0.92	+0.02	.	0.91	2 expl. var.
5	−1.05	.	+0.175	0.91	
6	−1.04	−0.02	+0.177	0.91	3 expl. var.

Source : [9] Table 1b, and figures on z'' kindly supplied by the authors.

These results may be interpreted so as to attach the main role in the explanation to *income*, with a clearly *negative* influence. The influence of the two education variables is secondary, with that of the level of education uncertain as to its algebraic sign, whereas inequality of education shows a positive influence. If income y' can be considered as a demand indicator for high qualification, its influence should be positive. This interpretation leads to a rejection of the demand–supply theory. But I have some doubts, already announced, whether the geographical units are not too small. A group of typically commuter municipalities, whose commuters work in the nearby large cities of Amsterdam, Rotterdam and The

Hague, do not reflect the demand for the commuters' qualifications. The municipalities happen to have high incomes and at the same time low inequality of incomes. In the next section we will find that for the larger units, the provinces, a completely different situation prevails.

3.4.5. Further research on the Netherlands

In an attempt to test the demand–supply theory I tried to construct a slightly more precise indicator for demand. From the American 1960 Census of Population quoted in [58] the percentage of manpower with higher education was found for the four main sectors: agriculture, manufacturing, trade and transportation, and services (defined as the remainder). For each of the Dutch provinces the total number of persons active in the four main sectors are known from the Dutch 1960 Census of Population. Multiplying the percentage with the higher education needed, as taken from American figures, a (probably overestimated) *index of demand* was derived. On the supply side, two indicators were used, in order to open up the possibility of different weights being given to manpower with secondary and to manpower with third-level higher education. At the same time it was assumed that the private cost of third-level education is related to income foregone, to be represented by a constant, reflecting the income of people with only secondary education.

The demand–supply theory was given a shape better adapted to the data available. As the variable representing income inequality we considered the upper decile income divided by average income (in Lydall's [42] notation P_{10}). Demand for and supply of people with higher education were represented by $d_1 + d_2$ and $s_1 + s_2$, respectively, where the indices 1 and 2 represent two subgroups: group 2 being university graduates and group 1 representing all other people with higher education. As set out in Section 2, the differences between demand and supply were taken as two explanatory variables, but the possibility was kept open that the weights of the two differences

$d_1 - s_1$ and $d_2 - s_2$ could be different: a scarcity in category 2 may be more important to explain inequality than the same scarcity in category 1. Taking into account that in the absence of inequality X'' must be 1 and that our method of calculating quantities demanded is based on American figures, a formula of the following shape was tested:

$$X'' = \xi_1(d_1 - s_1) + \xi_2(d_2 - s_2) + 1 + c, \qquad (3.4)$$

where c indicates the correction for the use of American figures. The data available do not permit us to introduce d_1 and d_2 separately, however. For this reason we combine $\xi_1 d_1 + \xi_2 d_2$ to $\xi Y''$ and specify the correction term c to be $\xi(\bar{Y}'' - \bar{Y}_o'')$, where the suffix o refers to the United States. Replacing s_1 and s_2 by $Z''' - U''$ and U'' (cf. Table 3.IV), respectively, we finally obtain, for the purpose of testing the demand–supply theory,

$$X'' = \xi Y'' - \xi_1(Z''' - U'') - \xi_2 U'' + 1 + \xi(\bar{Y}'' - Y_o'').$$

Our best result obtained runs

$$X'' = 1.21Y'' - 0.08Z''' - 1.16U'' - 11.4$$
$$(\mathrm{R} = 0.96). \qquad (3.5)$$

This is equivalent to putting $\xi = 1.21$; $\xi_1 = 0.08$ and $\xi_2 = 1.24$. This would leave us with an estimate of $\bar{Y}'' - \bar{Y}_o'' = -10.3$. The direct estimate of the percentage of active population with higher education in both countries yields

$$\bar{Y}'' = 10.4; \qquad \bar{Y}_o'' = 19.1,$$

implying a value for $\bar{Y}'' - Y_o'' = -8.7$. In order to test the stability of the regression coefficients found, we constructed Table 3.VIII, comparable with Tables 3.V, 3.VI and 3.VII, using normalized variables.

The negative influence of the supply variables and the positive influence of the demand variable is confirmed by cases 4 and 5.

In order to compare these results with those for the two

Table 3.VIII
Regression and multiple correlation coefficients R found for different combinations of variables explaining income inequality x''.

No.	Regression coefficients for			R	
	y''	z'''	u''		
1	0.84	.	.	0.84	1 expl. var.
2	.	0.81	.	0.81	
3	.	.	0.70	0.70	
4	1.03	−0.20	.	0.845	2 expl. var.
5	2.50	.	−1.72	0.95	
6	2.95	−0.42	−1.75	0.96	3 expl. var.

Table 3.IX
Regression and multiple correlation coefficients R found for different combinations of explanatory variables explaining x''.

No.	Regression coefficients for			R
	y	z'''	u''	
1	0.88	.	.	0.88
2	.	0.81	.	0.81
3	.	.	0.70	0.70
4	0.92	−0.04	.	0.88
5	1.02	.	−0.17	0.88
6	0.89	+0.27	−0.31	0.89

Table 3.X
Regression and multiple correlation coefficients R found for different combinations of variables explaining x'.

No.	Regression coefficients for			R
	y''	z'''	u''	
1	0.92	.	.	0.92
2	.	0.89	.	0.89
3	.	.	0.90	0.90
4	0.87	−0.055	.	0.92
5	0.91	.	+0.092	0.92
6	0.89	−0.054	+0.083	0.92

other countries and those obtained by Burns and Frech for the Netherlands (based on municipalities), we constructed similar tables for a few alternative variables, using y' instead of y'' (closer to Chiswick's material) for Table 3.IX and x' instead of x'' (Burns and Frech) for Table 3.X.

The results presented in the last two tables are less satisfactory than those of Table 3.VIII: the multiple correlation coefficients are lower and the supply influences are small and uncertain.

3.4.6. Some preliminary conclusions

The only case, in the present section, where a considerable influence of the level and the inequality of education on income distribution is found, is equation (3.5). In order to reduce income inequality, as measured by the highest decile divided by average income, to half of its 1960 value, that is, in order to attain $\Delta X = -2$, we need $\Delta U'' = 2/1.24 = 1.61$, meaning that the percentage of the population with university education should be more than doubled in comparison to the 1960 situation, when it was 1.4 per cent. Such favourable results were found in several other cases reported in Chapter 2; but most of the present results are much less favourable in that sense. From the various versions of the relationship found for the Netherlands one may wonder whether perhaps the use of the demand indicator as defined in Section 3.4.5 might not change the American and Canadian figures so as to show a stronger influence of education level or distribution on income inequality.* This will be undertaken in Section 3.5.

Another conclusion seems to be that municipalities are too small as units to compare to each other because of the different 'location' of demand and supply in our sense.

In a last attempt to compare our cross-section analyses we

* It is also conceivable that a longer-term influence on income distribution may be implicit in the influence of variable V, as already observed in Section 5. This is a suggestion made to me by J. P. Pronk and substantiated for Norwegian samples by Soltow [60].

Table 3.XI
Regression coefficients and R found in six cases.[a]

Case	R	Regression coefficient for				Country
		y	z	u	v	
A	0.96	2.95	−0.42	−1.75	.	Netherlands
B	0.94	1.02	−0.33	+0.315	−1.52	United States
C	0.92	0.88	−0.054	+0.083	.	Netherlands (provinces)
D	0.91	−1.04	−0.02	+0.177	.	Netherlands (municipalities)
E	0.89	0.89	+0.27	−0.31	.	Netherlands (provinces)
F	0.86	0.10	+0.27	+0.92	−1.61	Canada (provinces)

[a] Primes used to distinguish variables in Table 3.IV have been omitted in this table.

collect our 'best' cases from the various tables in the order of goodness of fit (Table 3.XI).

There are some regularities in this table worth mentioning. With the exception of case D, which we rejected because of the use of too small geographical units, the coefficients for y (or substitutes) fall from case A to F and so do (even including case D) the negative coefficients for z (or substitutes). Where available the v, representing other factors making for quality, exert considerable influence. This is an argument in favour of introducing such additional variables, as done by Chiswick in an inventive way.

3.5. Cross-section tests, parts of countries, continued

In this section the yardsticks for income inequality and for demand factors will be changed in a further attempt to explore Chiswick's material for the United States. To begin with, the variables used were:

X' = $P(10)$ in Lydall's notation; that is, the upper decile divided by the median of income by state;

Y' = a demand index for third-level educated people, equal to the average ratios of the active population with higher education in the four large sectors (agriculture, manufacturing, trade and transportation, and other services) weighted with the percentages these sectors have among the active population of the state considered;

Z = average years of schooling;

U' = standard deviation (in years of schooling).

The following results were obtained:*

$$X' = 0.047Y' - 0.194Z + 0.160U' + \text{constant} \quad (\text{R} = 0.77). \tag{3.6}$$

Since $\bar{X}' = 2.28$, and perfect income equality would yield $X' = 1$, halving inequality would require $\Delta X' = -0.64$. This can be obtained, for instance, by halving U', or $\Delta U' = -1.9$, and by changing Z, $\Delta Z = 1.75$, bringing it from 10.2 to 12 years. *Such values seem to be attainable without undue effort.*

Alternatively, using the additional explanatory variable V, income at zero schooling, we obtain

$$X' = 0.029Y' - 0.112Z + 0.21U' + \text{constant} \quad (\text{R} = 0.83),$$

from which we infer that income inequality can be halved again by halving U' and, in addition, raising Z from 10.2 to 12.4 years, again without undue effort.

Finally, for the United States three variables were calculated representing excess supply for the three levels of schooling and indicating the difference between supply and demand as calculated before. They are designated T_1, T_2 and T_3, expressed as percentages of the total labour force, and are now used as explanatory variables. The result obtained is

$$X' = 0.0008T_1 - 0.0326T_2 - 0.015T_3 + 2.0 \quad (\text{R} = 0.73), \tag{3.7}$$

Since for the United States as a whole $T_1 + T_2 + T_3 = 0$ and assuming that $\Delta T_3 = 0.3\Delta T_2$, we can estimate the values of the explanatory variables that again would halve X', that is, make $\Delta X' = -0.64$. The resulting values are compared with the observed values for 1960 as follows:

* F. Wim Blase performed the calculations.

	T_1	T_2	T_3
Observed	12.6	−7.9	−1.1
Required	−9.2	+8.9	+3.9

In order to judge the degree of realism of the required figures, we have to know the observed values of supply, as percentages of the active population, of the three categories. In 1960, there were around 40 per cent for the first and second level of education and 20 per cent for the third, implying that the changes required do not seem to be out of reach. They do imply the necessity of creating a shortage of people looking for jobs requiring primary education (according to 1960 views) and a surplus of people looking for jobs previously requiring third-level education in order to raise the lowest and to depress the highest incomes.

Another way of judging the feasibility of our new values for T_1, T_2 and T_3 is to analyse the actual 1960 figures for individual states. Thus, while the average for T_1 amounted to 12.6, as already observed, we find seven states where it was ≤ 5; and while the average for T_3 amounted to -1.1, in fourteen states the figure was $\geq +1$.

Our final view on the prospects for income distribution in the United States will be given in Chapter 6, however.

3.6. An incomplete absolute price equation

Finally, a few examples may be presented of an absolute price equation, that is, a price equation for incomes of different groups instead of income distribution parameters. The equations to be presented are incomplete in the sense that they do not include demand factor variables, because either all groups considered are supposed to live under the same demand conditions or no data are available on possibly differing demand conditions for the groups considered.

* I am indebted to Jaap Jansen and Hans Opdam for having performed the computations.

The first and most important example is based on American material taken from Fuchs, which I obtained by the courtesy of Professor Mincer (cf. also [45]).

It supplies us with the average hourly earnings of non-agricultural white males in the USA in 1959, for seven age groups and six groups for the years of completed schooling. We specify that in the following equations x_1 means age, x_2 schooling (both in years), and y earnings in cents per hour. Of course the linearity of the relation is far from certain and we will introduce alternative mathematical expressions. But it appears that already the linear relationship yields a high correlation coefficient; for the complete material we obtain

$$y = \underset{(0.47)}{3.48x_1} + \underset{(1.68)}{14.5x_2} - 17.2 \qquad (\mathrm{R} = 0.88), \tag{3.8}$$

and excluding the highest age group (65 and over) even

$$y = \underset{(0.55)}{4.36x_1} + \underset{(1.69)}{13.4x_2} - 30.7 \qquad (\mathrm{R} = 0.90). \tag{3.9}$$

Figures in parentheses are standard deviations.

A look at the scatter diagram suggests that a curvilinear relationship with respect to both independent variables will significantly improve these results and hence the following two relations were also estimated for the complete material,

$$y = \underset{(2.55)}{13.0x_1} + \underset{(1.45)}{14.3x_2} - \underset{(0.030)}{0.115x_1^2} - \underset{(48)}{176} \quad (\overline{\mathrm{R}} = 0.91), \tag{3.10}$$

and

$$y = \underset{(1.42)}{9.85x_1} - \underset{(3.6)}{16.5x_2} + \underset{(0.17)}{0.88x_2^2} + \underset{(0.046)}{0.35x_1x_2} - \underset{(0.017)}{0.120x_1^2} + \underset{(32.7)}{15.3} \quad (\overline{\mathrm{R}} = 0.97). \tag{3.11}$$

It will be observed that all regression coefficients are significant at the 1 % level, except the intercept. An interesting feature of equation (3.11) is the term x_1x_2 which shows the influence of '*combined scarcities*' of ability and experience.

Table 3.XII
Actual and calculated (in parentheses) values of hourly earnings (cents) of non-agricultural white males with different schooling and age; USA, 1959.

Age	Years of schooling completed					
	2	6.5	10	12	14	17
16	98 (124)	151 (109)	145 (122)	140 (139)	161 (163)	. .
22	148 (160)	171 (154)	177 (174)	190 (196)	209 (224)	226 (280)
29	178 (191)	218 (196)	241 (225)	257 (251)	279 (284)	330 (347)
39	200 (215)	243 (236)	276 (277)	302 (310)	350 (350)	469 (424)
49	195 (215)	254 (252)	278 (305)	316 (345)	403 (392)	533 (476)
59	211 (192)	250 (244)	290 (309)	334 (356)	417 (410)	514 (505)
66	190 (160)	226 (224)	284 (298)	372 (350)	362 (409)	515 (511)

Table 3.XII shows the observed and the estimated values of y for all combinations of x_1 and x_2 (for which the middle values of the intervals have been taken).

The relationship found can also be transformed into one using as independent variables: schooling and work experience $x_3 = x_1 - x_2 - 6$,

$$y = -6.0x_2 + 1.11x_2^2 + 8.4x_3 + 0.11x_2x_3 - 0.12x_3^2 + 70. \tag{3.12}$$

From (3.12) it appears that increases in x_2, schooling, are paid an increasing income differential, while increases in x_3, experience, are paid a decreasing income differential. Experience is paid a maximum for 30 to 40 years, depending slightly on schooling.

Another, extremely simple, example consists of an estimate based on income and schooling differences between the provinces of the Netherlands. Here a relation was obtained

$$\underset{(0.09)}{y = 1.37x_2} - 6.3 \qquad (R = 0.92), \tag{3.13}$$

where y was measured in thousands of guilders (1960) and x_2 in years of schooling completed. The data are those published by the Dutch Central Bureau of Statistics [12].

A number of other studies concerning the impact of personal characteristics on incomes earned have been made by other authors: especially by those belonging to the human capital school, such as Mincer [45] or Chiswick [16]; by authors using samples of individuals, many of whom have been quoted and interpreted by Jencks and his collaborators [33]; and by De Wolff and Van Slijpe [20]. I exclude from this list research on income distribution since most of its results cannot be used directly to get an insight into the question I want to consider as the central issue, namely what is the influence of a number of personal characteristics on income proper (and not on their distribution).

A striking feature of a comparison between the various studies is a considerable difference between inquiries dealing with *groups* of people as their units of observation and inquiries dealing with *individuals* as their units. As a rule much higher correlations are obtained by the former than by the latter. The results given above all show correlation coefficients above 0.9, whereas the results of work with individual data show correlation coefficients of at most 0.7, meaning that at best half of the variance can be explained. This applies to De Wolff and Van Slijpe who, from Husén's [31] material, derive the relationship explaining income with the aid of social class of parents, IQ and years of schooling. Much lower correlations are found, as a rule, by Bowles [5] who arrives at 0.40; in this correlation coefficient a minor contribution is obtained from schooling. Similarly, the correlations found by Chiswick [16] for the states of the USA between the natural logarithm in earnings and years of schooling show a correlation coefficient between 0.33 and 0.57. A possible explanation of this discrepancy is, as mentioned also by Jencks [33], that a number of other factors interfere, such as *non-cognitive skills*, *demand* factors and just *luck*. The importance of non-cognitive skills may be illustrated by an attempt to explain income differences in the Netherlands by differences in degree of independence. These, even though measured in a very crude way, alone show an $r = 0.89$, again for 21 groups (Chapter 4, Table 4.II).

Because of my adherence to a demand–supply theory I also submit that in addition to personal characteristics demand factors co-determine an individual's income. The neglected supply (non-cognitive) and demand factors evidently do not correlate strongly with schooling; otherwise the correlation coefficients for inquiries with individuals would have been higher. But they must cancel each other out a good deal, otherwise the group correlations with schooling would not be so high. This latter assumption may be tested by comparing the values of the regression coefficients found in some inquiries with groups with the values found from inquiries with individuals.

The influence of one more year of schooling on incomes expressed in dollars per year can be derived from the linear equations (3.8) and (3.9) by multiplying the influence on hourly earnings by the number of working hours per year divided by 100, amounting roughly to 20. The results are 268 to 290 if we take the regression coefficients of x_2, and 356 to 360 if we replace x_1 by the expression $x_1 = x_3 + x_2 + 6$, x_3 representing years of work experience.

From the results obtained by De Wolff and Van Slijpe, using the coefficients in front of Z_4 (which the authors report to be the most important) in equations 16 and 18, respectively, we find for one more year of schooling an income difference of 1700 respectively 1400 Swedish kronor, or about \$ 340 and \$ 280.

From the analysis using the provinces of the Netherlands, equation (3.13), we find Dfl. 1370 or \$ 380.

While it is significant, on the one hand, that the orders of magnitude are the same, it is also significant that the American figures are lower than the Dutch figure. Keeping in mind that the Swedish figures refer to people of age 35, we find that the corresponding figure for the USA is lower than the Swedish figure too. This reflects a lower scarcity of more qualified people in the USA than in Sweden and than in the Netherlands.

Chiswick [16] finds that this same difference is shown by the negative correlation (-0.79) between (1) the regression coefficients of $\ln E$ (natural logarithm of earnings) on years of schooling and (2) the states' income (averages of natural logarithms [16] Table 3-3). In addition the regression coefficient is higher (0.14) for Puerto Rico than for the USA (0.11), [16] Table 4-2. Support for the thesis that income differences between groups with different schooling correlate inversely with their scarcity can also be found in historical comparisons as made for the period 1900–1963 in the USA by Ullman [74] Table 3.

So far we have assumed that relative scarcity is mainly changed by changes in supply. Historically these changes have been considerable and over a long period such as the one considered by Ullman [74], 1900–1963, they are largely responsible for the reduction in income inequality. There are also changes in demand, however, which affect relative scarcity. In Chapter 6, I have tried to disentangle these two influences and found that in fact it is a 'race' between supply (by education) and demand (by technological development) which determines the changes in relative scarcity of any type of manpower.

In conclusion we may state that for large groups of individuals with different schooling and different length of work experience their income averages differ significantly. Correlations obtained for such group characteristics and their incomes are above 0.9 in the cases analysed statistically, covering the United States (1959) and the Netherlands (1960). The lower correlations found in inquiries with individuals need not be ascribed to random factors or 'luck' only; there are other personal characteristics which affect incomes, such as independence, health and family size. In addition, demand factors play a role; and relative scarcity is the more important explanation of income differences.

This implies that the income increase to be obtained by one more year of education can be read from our formulae for only small numbers of persons who make a decision of whether or not to increase schooling. The additional income obtainable is not a fixed amount for all countries or all times; we found it to be higher, the lower the average income of the countries considered.

Among the inquiries made with single persons, there remains, for the time being, an as yet unsolved contrast between, for instance, Bowles who finds a very limited influence of schooling and results obtained by De Wolff and Van Slijpe and the present author who find a significant influence. These are not necessarily due to the choice of the other explanatory variables used in the regression equations.