THE DISTRIBUTION OF LIFETIME EARNINGS *

J.H.M. NELISSEN

Tilburg University, 5000 LE Tilburg, The Netherlands

Assuming that earnings at a particular age are distributed log-normally, it is shown that the distribution of lifetime earnings can be simulated. For that purpose estimations for the course of μ_t and σ_t^2 are made.

1. Introduction

Little is known about the real distribution of lifetime earnings. Generally, the distribution of lifetime earnings is derived from panel data or a pooling of cohort and cross-section data, covering only a relatively small part of the period during which a person earns money. In the most favourable cases the period under investigation spans about ten years [e.g., Creedy and Hart (1979), and Fase (1969)]. Starting from the widely accepted assumption that at a particular age, the earnings are distributed lognormally, the development of the mean and variance of the logarithms of earnings to age have been supposed to follow a quadratic and a linear function in the variable age, respectively. The functions are estimated, and this way the distribution of lifetime earnings has been determined. In a similar way it is possible to describe the distribution of earnings in a particular year. Besides that these functions for different cohorts are based on (too) few observations, there is another important disadvantage. It is hardly possible to derive the distribution of earnings for the future or for the past. For, the parameters of the aforementioned quadratic and linear functions for the mean μ_t and variance σ_t^2 differ between generations and years. This is, of course, because these functions are not based on theoretical arguments, but only on empirical insights, although human capital theory, for example, is able to explain the quadratic form of μ_t. Our purpose now is to estimate μ_t and σ_t^2 using a database covering an unusually long period and including as many theoretical considerations as possible.

2. The data

Since 1962, the Netherlands Central Bureau of Statistics (henceforth, NCBS) provides us with (aggregated) earnings data for individuals. [1] We use primary income, excluding income from capital,

---

[1] Before 1962 the household was the income unit, so that in the case of married people, each individual's income was not available.

as the income concept. This means that we take the labour income of the self-employed, employees, and managers of private and limited liability companies and corporations. [2] From the NCBS (1967a, b, 1972a, b, 1977, 1979, 1980, 1982, 1984 and 1985) we have data for different age groups for the years 1962, 1967, 1970, 1975, 1979 and 1981. These data are transformed into constant prices of 1980.

From these pooled cross-section data we can derive 53 observations for the mean income, and 31 for its variance.

3. The model

Our point of departure is the statistical model, adopted from Aitchison and Brown (1957) and Fase (1969), which describes the distribution of the individual income by age. In this, the individual income has been considered as a random variable related to age. The individual is assumed to start his career at age s, at which he or she receives an annual income equal to y_s. This variable y_s is assumed to follow a two-parameter lognormal distribution and age is considered to be a discrete variable. If we also assume that the distribution of the proportional changes in income with increasing age is a normal one, then income follows a lognormal distribution for all ages [see Hart (1973)]. Thus,

$$P(\mathbf{y}_t \le y_t) = \wedge \left(y_t;\ \mu_t,\ \sigma_t^2 \right),\tag{1}$$

with

$$\wedge \left(y_t;\ \mu_t,\ \sigma_t^2 \right) = \int_0^{y_t} \left(y \cdot \sigma_t \cdot \sqrt{2\pi} \right) - 1 * \exp - \tfrac{1}{2}\left[(\log Y - \mu_t)/\sigma_t \right]^2 \, \mathrm{d}y.$$

The usual procedure is to determine the distribution of the complete age–income profile by introducing a quadratic function in the variable age for μ_t and a linear one for σ_t^2, whereby in some cases dummy variables for different cohorts are also used [e.g., Creedy and Hart (1979)]. We follow another method by introducing theoretical interpretable variables.

For the variable μ_t we use the following variables:

(i) The development in the logarithm of national income, corrected for the working population in proportion to the total population, *NI*.
(ii) Because female labour force participation rates differ from male ones, and because of their development in the course of time, these labour force participation rates (corrected for part-time participation), *LFF*, are inserted.
(iii) A higher level of education of the population under consideration can influence the proportion of national income which the working population receives. Therefore, the proportion of people with secondary education, *EDS*, has been used.
(iv) Unemployment can have the opposite effect, which argues for the insertion of the unemployment rate, *UN*. This variable could also correct for the employment benefits (see footnote 2).

---

[2] Some benefits for the unemployed are included and cannot be isolated. However, their impact during the period under investiagation was limited.

(v) Human capital theory shows a relationship between income and age. We insert this by using dummy variables for different age groups, *AGExx*. [3]

The difference with the afore-mentioned method is that the coefficients of *AGExx* are constant in time and that changes in the income–age profile are the consequence of changes in other variables. The GLS estimation result is (standard errors are shown in parentheses) [4]

$$\log \alpha = 11.281 - 0.145 \ CT2 + 0.653 \ NI - 0.967 \ LFF - 0.017 \ UN$$
$$\quad (0.286) \ (0.043) \qquad (0.071) \quad (0.172) \qquad (0.006)$$

$$+ 1.612 \ EDS + 1.265 \ AGE20 + 1.624 \ AGE30 + 1.777 \ AGE40$$
$$\quad (0.472) \qquad (0.218) \qquad\qquad (0.167) \qquad\qquad (0.163)$$

$$+ 1.785 \ AGE50 + 1.713 \ AGE60 + 1.109 \ AGE70$$
$$\quad (0.166) \qquad\qquad (0.165) \qquad\qquad (0.259) \tag{2}$$

$$R^2 \text{adj} + + 0.986, \qquad \sigma_\epsilon = 0.064.$$

Here, an extra dummy variable *CT2*, which equals one for observation before 1971, has been inserted to correct for a shift in mean income after 1970. The goodness of fit, as measured by R^2adj, is very high. All variables are significant at the 0.05 level and have the expected sign.

This equation has been used to simulate the mean of the age–earnings profile of individuals for the year 1950. The result is given in the second column of table 1. The third column shows the average household income for that year. Data on the mean individual income are not available. Using female labour force participation rates we have made a correction for this, as given in the last column. Opposite the data used for the estimation (and consequently for the simulation), the 1950 data are not adjusted for persons who have a job during only part of the year (school-leavers,

Table 1
Simulation of the mean age–earnings profile for the year 1950 (in constant prices of 1980).

Age	Model	Income *hh* [a]	Idem, corrected
< 20	5000	4447	4447
20–24	9100	8489	7900
25–29	14100	13600	12600
30–34	16200	18853	16500
35–39	17500		
40–44	18600	23089	19600
45–49	18800		
50–54	18600	23084	20000
55–59	18300		
60–64	16200	20837	18900

[a] *Source:* NCBS (1954).

[3] For age 30, for example, *AGE30* takes the value of 1, whereas for age 27, *AGE20* takes the value of 0.3 and *AGE30* a value of 0.7.

[4] The dependant variable is not the mean of the logarithm of income, μ_t, but the logarithm of the mean of the income, $\log \alpha_t$, which can be transformed to μ_t via $\mu_t = \log \alpha_t - 0.5 \ \sigma_t^2$. The reason for this is that we have more observations for α_t.

vacation workers, and so on) and they do include capital income of workers. This means that observed income has been underestimated in the younger age groups and overestimated in the older age groups. Taking this into account, we are allowed to conclude that the simulation results fit very well. We also made a number of predictions for the year 2000 (for different growth scenarios) which led to very plausible results [see Nelissen (1988)].

In a similar way a regression equation for the coefficient for variation σ_t/μ_t has been estimated. Here, the dummy variables for age, *AGExx*, are inserted to cover variance increasing elements with growing age. Economic growth is assumed to diminish income inequality. Therefore, the variable *NI* is included in the equation. The same holds true for a better educated population. This is embodied in the variable *EDS*. Unemployment is expected to raise income inequality and the variable *LFF* is used for the effect of female labour force participation. The sign of *LFF* is not known beforehand. Theory is not unequivocal. Also, the labour force as a proportion of the total population, *LF*, is inserted in the regression equation. A relatively large working population could diminish income inequality, because the possibility for shortages in labour force are minor. The regression result is

$$\sigma/\mu = 3.271 - 13.079 \, LF - 0.477 \, NI + 0.149 \, LFF + 0.048 \, UN$$

$$\quad (1.172) \quad (3.573) \quad\quad (0.064) \quad\quad (0.093) \quad\quad\;\; (0.012)$$

$$\quad - 5.737 \, EDS - 0.079 \, AGE30 - 0.093 \, AGE40 - 0.062 \, AGE50, \tag{3}$$

$$\quad\;\;\; (3.464) \quad\quad\;\; (0.033) \quad\quad\;\;\; (0.027) \quad\quad\;\;\; (0.031)$$

$$R^2 = 0.894, \qquad \sigma_\epsilon = 0.041.$$

The adjusted R^2 is rather high, all coefficients are significant and they have the expected sign. In the same way as we did for μ_t, it is now possible to derive the coefficient of variation for, for example, 1950 and 2000.

The complete distribution of the age–earnings profile up to age x for a specific cohort or year can now be derived, after solving μ_t and σ_t for each t from eqs. (2) and (3), by means of convolution of the distributions for each age (taking account of the probability that an individual will survive and participate in earning money at that age) or by simulation. Given the analytical calculation problems inherent to the first method, the simulation approach seems to be preferable. In Nelissen (1988), the income distribution is implicitly derived within a micro-simulation context.

4. Conclusion

Starting from the statistical model used by Aitchison and Brown (1957) and Fase (1969) and using an explanatory model for μ_t and σ_t/μ_t it is possible to derive the distribution of lifetime income of a cohort or the distribution of income in a particular year in the future.

References

Aitchison, J. and J.A.C. Brown, 1957, The lognormal distribution (Cambridge University Press, Cambridge).
Creedy, J. and P.E. Hart, 1979, Age and the distribution of earnings, Economic Journal 89, 280–293.
Fase, M.M.G., 1969, An econometric model of age–income profiles, Ph.D. dissertation (Universitaire Pers Rotterdam, Rotterdam).

Hart, P.E., 1973, The comparative statics and dynamics of income distribution, Discussion paper, series A, no. 49 (University of Reading, Reading).

NCBS, 1954, Inkomensverdeling 1950; Aanvullende gegevens (de Haan, Utrecht).

NCBS, 1967a, Inkomensverdeling 1962 en vermogensverdeling 1963 (Staatsuitgeverij, The Hague).

NCBS, 1967b, Inkomensverdeling van werkende gehuwde vrouwen 1962 (Staatsuitgeverij, The Hague).

NCBS, 1972a, Inkomensverdeling 1967 en vermogensverdeling 1968 (Staatsuitgeverij, The Hague).

NCBS, 1972b, Inkomensverdeling van werkende gehuwde vrouwen 1967 (Staatsuitgeverij, The Hague).

NCBS, 1977, Inkomensverdeling 1970 en vermogensverdeling 1971 (Staatsuitgeverij, The Hague).

NCBS, 1979, Inkomensverdeling van werkende gehuwde vrouwen 1970 en 1973 (Staatsuitgeverij, The Hague).

NCBS, 1980, De personele inkomensverdeling 1975, deel I (Staatsuitgeverij, The Hague).

NCBS, 1982, De personele inkomensverdeling, Werkende gehuwde vrouwen 1975 en 1976 (Staatsuitgeverij, The Hague).

NCBS, 1984, De personele inkomensverdeling 1979 (Staatsuitgeverij, The Hague).

NCBS, 1985, The personele inkomensverdeling 1981, Individuen (Staatsuitgeverij, The Hague).

Nelissen, J.H.M., 1988, A model for the age-earnings profile, Working paper (Tilburg University, Tilburg) forthcoming.

Nelissen, J.H.M., 1989, The redistributive impact of social security settlements on lifetime income, Ph.D. dissertation, forthcoming.