Appendix G: Estimation of the Mean Income for the Open-Ended Class

Wolfle's cross tabulations contain the number of people in the following annual-income classes:

1   Less than $1,500

2   $1,500 to $2,500

3   $2,500 to $3,500

4   $3,500 to $4,500

5   $4,500 to $5,500

6   $5,500 to $6,500

7   $6,500 to $8,500

8   $8,500 to $12,000

9   Over $12,000

For the first eight groups, we used the midpoint of the class as the average income in the interval.[1] In order to estimate the mean in the ninth class, we assumed that the upper tail of the income distribution follows the Pareto distribution. The validity of this assumption has been discussed in such studies as Klein (1962). Other investigators of the returns to education have estimated the mean of the open-ended class by using the Pareto distribution (see, for example, Becker, 1964). If the tail of the distribution is Pareto, the mean of the open-ended class ($\overline{Y}_n$) can be expressed as

$$\overline{Y}_n = Y_L \frac{\alpha}{1 - \alpha}$$

[1] We used $750 for the first class.

where Y_L is the lower limit of the open-ended class (that is, $12,000 in this instance), and where α is estimated from the data. Since the Pareto is relevant only in the upper tail of the income distribution, we made two estimates, assuming first that the tail began in the $4,500 to $5,500 class, and then that it began in the $5,500 to $6,500 interval. For all three states (Illinois, Minnesota, and New York), only the professional, semiprofessional, and sales occupations contained people in the open-ended category.[2] Because of limitations in sample size, one estimate of α was made for all three states combined for each of the sales and semiprofessional occupations. A separate α was estimated in each of the three areas for professionals.

TABLE C-1
Estimated α and mean income of males for the open-ended class

	α_1	$\overline{Y}_1$	α_2	$\overline{Y}_2$
Professional, Minnesota	2.3	$ 21,000	2.2	$22,000
Professional, Illinois	2.2	22,200	1.7	29,000
Professional, Rochester, N.Y.	1.01	1,212,000	1.6	32,000
Semiprofessional	2.4	20,600	2.2	22,000
Sales	2.0	24,000	2.3	21,000

The estimates obtained when the upper tail was assumed to begin with the $4,500 to $5,500 class are denoted by the subscript 1; the estimates based on the data beginning with the $5,500 to $6,500 interval are denoted by the subscript 2. Except for professionals from Rochester, New York, the differences implied by the estimates are not large. In our computations we used the estimates given in the $\overline{Y}_2$ column.

In some calculations, we required the mean income for people in all occupations. We combined the above data to obtain an estimate of $25,000, which was used for all three areas. We also made some computations based on $20,000. The regressions reported in Chapter 3 indicate that the use of $20,000 rather than $25,000 did not result in substantial changes in the coefficients.

---

[2]In fact, for Rochester, New York, and Illinois nearly all cases were for professionals.