177-73

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

AGE, PERIOD, COHORT, AND EDUCATION EFFECTS ON EARNINGS BY RACE--AN EXPERIMENT WITH A SEQUENCE OF CROSS-SECTIONAL SURVEYS

H.H. Winsborough

UNIVERSITY OF WISCONSIN-MADISON

AGE, PERIOD, COHORT, AND EDUCATION EFFECTS ON EARNINGS BY RACE--AN EXPERIMENT WITH A SEQUENCE OF CROSS-SECTIONAL SURVEYS*

H.H. Winsborough

October, 1973

* Forthcoming in Social Indicator Models, Kenneth C. Land and Seymour Spilerman (eds.), New York: Russell Sage Foundation.

ACKNOWLEDGMENTS

The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. The conclusions are the sole responsibility of the author.

Certain data used in this analysis were derived by the author from a computer tape file furnished under the joint project sponsored by the U. S. Bureau of the Census and the Population Council and containing selected 1960 Census information for a 0.1 percent sample of the population of the United States. Neither the Census Bureau nor the Population Council assumes any responsibility for the validity of any of the figures or interpretations of the figures presented herein or based on this material.

Data for the 1962 Occupational Change in a Generation Survey was kindly provided by Professors David Featherman and Robert Hauser whose purchase of that data was supported by National Science Foundation Grant GI31604X.

Current Population Survey person files for 1968 and 1970 were kindly provided by Mr. Al Rosenthal of the Rand Corporation.

Much of the computing for this paper was accomplished on the computing facility of the Center for Demography and Ecology, University of Wisconsin, Madison, which is supported by a "Centers Grant" from the National Institute of Child Health and Human Development, Grant Number 1-P01-HD058F6-01A1.

The assistance of Peter Dickinson in computational work is gratefully acknowledged. Careful criticism of a first draft of this paper by I. Richard Savage, Stanley Masters, and Kenneth Land have been most helpful in preparing a final draft as have discussions with Otis Dudley Duncan, Karen Mason, and Charles Palit. Errors remaining in the paper probably derive from the author's decision not to take all of their good advice.

Age, Period, Cohort, and Education Effects on Earnings by Race--

An Experiment with a Sequence of Cross-Sectional Surveys

ABSTRACT

This paper does three things simultaneously.

First, it argues that the Census Bureau should release basic records from current population surveys taken in the past and in the future because they constitute a unique resource for research into the dynamics of poverty. The argument is made by negative example. An assortment of available files are used to estimate an earnings model requiring a series of cross-sectional surveys. The example shows that important information about the nature, causes, and consequences of poverty can be derived from such models but that period differences in the source of the data becloud interpretation in unfortunate ways.

Second, the paper poses a simple human capital model and, in the context of the above experiment, compares parameter estimates for blacks and whites. It finds very similar patterns of response to background factors influencing earnings in the two races. By implication it is clear that the inferior earnings of blacks are <u>not</u> to be improved by more schooling, by better schools, nor by a more favorable pattern of work experience.

Finally, the paper accomplishes its two above-described tasks by exercising an estimation method which permits separation of age, period, and birth cohort effects. The results document the existence of a separate cohort effect on earnings. People born into relatively large cohorts earn, <u>ceterus paribus</u>, about 80 percent of the wages of persons born into small cohorts. This finding points out a less than obvious stake poverty policy makers have in the creation of population policy. Any population policy which permits or encourages cyclic fluctuation in birth rates--instant zero growth is one example--is likely to engender waves of poverty moving through the age distribution over time.

AGE, PERIOD, COHORT, AND EDUCATION EFFECTS ON EARNINGS BY RACE -- AN EXPERIMENT WITH A SEQUENCE OF CROSS-SECTIONAL SURVEYS

THE CURRENT POPULATION SURVEY AS A DATA SOURCE FOR MODELS OF SOCIAL CHANGE

A large fraction of the presently available indicators of the status of and change in American society are derived from the Current Population Survey (CPS) conducted monthly by the Bureau of the Census. Originally conceived of as a mechanism for collecting unemployment statistics, this large, highly professional, and thoroughly routinized survey presently provides at least annual data on a whole host of social and economic variables.

It is my not-very-original contention that these surveys also provide a large and relatively untapped resource for the construction of models. Were basic CPS records for past and future data routinely made available to scholars, I believe a host of new and important social indicators would be forthcoming from the model building which would ensue. The argument of this paper, then, is that the CPS archive represents an important resource for constructing interesting models of social change -- models which can yield useful indicators as their by-product.

This argument will be made by example. I shall present a simple human capital model to explain individual earnings. Estimating the parameters of the model separately for blacks and whites should yield information on the changing sources of black-white income differences. This estimation requires tables of income by education, age, and race for several periods. Ideally, such tables should be constructed from a sequence of surveys conducted and processed in such a homogeneous way as to minimize "methods variance" between them. A sequence of annual CPS files from, say, 1960 to 1970 would be nearly ideal. Since these files are not presently in the public domain, I shall make-do with surrogate sources which are generally available and rather widely used. On the one hand, I hope this strategy will provide a convincing argument for the potential utility of a model which requires repeated cross-sectional data. At the same time, the strategy will display the ambiguities of interpretation which arise from the compounding of "methods" and "real" effects.

The outline of the paper is as follows: First, the model itself is discussed in order to make clear its data requirements. Second, the estimation of the model is discussed. Third, the data actually used to estimate parameters is described and some of the data's problems are noted. Fourth, the estimated parameters are presented and the results discussed. Fifth, ways in which using CPS files would have made interpretation more secure are considered. Finally, consideration is given to some of the problems which the author's work suggests would be encountered in a facility designed to routinely estimate parameters of social change models from CPS files.

THE MODEL

The usual human capital model presumes children enter school with some level of human capital. This capital is a combination of innate ability and nonschool training. Although the value of this capital varies over individuals, its mean in the aggregate should vary only by birth cohorts as secular trends in ability and preschool training are captured and capitalized upon by each new school entrance cohort. Let us call this value A_c where the subscript specifies the relevant cohort.

Society requires that this preschool capital be invested in education for a period of time. Here I presume that investment increases capital at a rate r_i for the i^{th} additional unit of schooling. Hence, the capital available on the termination of education after t years is $A_c \prod_{i=1}^{t} (1 + r_i)$.

On the termination of education, capital is invested in the labor market yielding income at some rate R during the first year. Ideally, the value of R is presumed constant over individuals but may vary by time period. Let us add a subscript p to R to indicate its variation by period. Thus, income in the first year of employment is given by:

$$I_{pct} = R_p A_c \prod_{i=1}^{t} (1 + r_i) .$$

(Of course, the value of R_p for a period might be different between the races if only limited kinds of jobs were available to blacks and the supply or demand for those jobs were different from that of the labor force as a whole.)

Subsequent to the first year of employment, two things are presumed to happen. First, human capital is increased by virtue of the experience of working. Second, capital is depreciated by aging, and by obsolescence of both training and prior experience. One would like to measure experience and depreciation separately and directly but such measures are not routinely available. The best that can be done with available data is to presume individuals have been working since completing school and to use that elapsed time in years as a measure of experience (Thurow, 1968:235).

Formally, e, the index of experience, can be estimated as $y - (t + 6)$ where y is years of age, t is the termination level of schooling, and 6 is an estimate of age on entrance to school. Clearly, such a measure would also involve the best available index of depreciation, i.e., year of age. The balance of the two effects on the quantity of capital can be computed as $\prod_{j=1}^{e} (1 + \rho_j)$ where ρ_j is the rate of increase or decrease in capital during the j^{th} year after leaving school and e indexes the number of years of experience.

Hence, our income model would become:

$$I_{pcte} = R_p A_c \prod_{i=1}^{t} (1 + r_i) \prod_{j=1}^{e} (1 + \rho_j) \quad .$$

In estimating parameters for models somewhat similar to this one, it is not uncommon to compute an experience index from the individual data. A simplification of the model makes this computation unnecessary and clarifies the interpretation. Suppose we think of computing the value of income from this equation for two individuals with different levels of education and, perhaps, ability, but of the same age in the same period. If the first individual has the minimum education for his group, his equation might be:

$$R\, A_1\, (1 + r_1)\, (1 + r_2) \ldots (1 + r_t)\, (1 + \rho_1)\, (1 + \rho_2) \ldots (1 + \rho_{e-1})\, (1 + \rho_e)$$

where A_1 indicates his ability level. If the second individual has one unit more of education (and hence one unit less of experience) his equation would be:

$$R\, A_2\, (1 + r_1)\, (1 + r_2) \ldots (1 + r_t)\, (1 + r_{t+1})\, (1 + \rho_1)\, (1 + \rho_2) \ldots (1 + \rho_{e-1})$$

where A_2 is his, perhaps greater, ability level. The ratio of the second income to the first is simply $\dfrac{A_2\, (1 + r_{t+1})}{A_1\, (1 + \rho_e)}$,

the degree to which the investment of an additional year in education improves human capital over the improvement available by investing the year in work experience multiplied by an ability selection factor.

In general, then, if we define the rate of return to education, call it r'_i , as the excess over returns to capital via experience and recognize that it includes ability selection, our model simplifies to :

$$(1) \qquad I_{pcyt} = R_p A_c \prod_{j=1}^{y} (1 + \rho_j) \prod_{i=1}^{t} (1 + r'_i)$$

where ρ_j is the return to the maximum level of experience of the year-of-age group y and A_c is the ability level of the earliest school terminators in cohort c.

ESTIMATION

The model of Equation (1) specifies the expected income of an individual of age y in the p^{th} period (and consequently in cohort c) having units of education t. Writing Y_y for $\prod_{j=1}^{y} (1 + \rho_j)$ and T_t for $\prod_{i=1}^{t} (1 + r'_j)$, Equation (1) becomes:

$$I_{pcyt} = R_p A_c Y_y T_t \quad .$$

Multiplying and dividing by the mean income for the several ages, periods, and education groups and observing that $\overline{I} = \overline{A}\ \overline{R}\ \overline{Y}\ \overline{T}$, where means are geometric, yields:

$$(2) \qquad I_{pcyt} = \overline{I} \frac{A_c}{\overline{A}} \frac{R_p}{\overline{R}} \frac{Y_y}{\overline{Y}} \frac{T_t}{\overline{T}} \quad .$$

Taking logs of both sides, using a caret to designate the log of the variable and adding a stochastic error term Equation (2) becomes:

$$(3) \qquad I'_{pcyt} = \overline{I}' + (A'_c - \overline{A}') + (R'_p - \overline{R}') + (Y' - \overline{Y}') + (T' - \overline{T}') + e.$$

Given tables of means, variances, and numbers of observations for log income by age, period, and education, estimation of parameters in this model becomes a problem in the estimation of effect-parameters in a three-way analysis of variance design with unequal cell sizes. The model presumes there exists "main effects" for age, period, and education as well as "interaction effects" for cohorts. For a given educational level, these interaction effects for cohorts can be thought of as the series of diagonal interactions in the age, period table. As is common in analysis of variance problems, the parameters are not estimable without setting some side conditions. We first set the usual side condition that the weighted sum of effect parameters for each single effect (e.g., age or period) is equal to zero where the weights are the number of observations. This side condition is not by itself sufficient to make the parameters of (3) estimable because of the linear constraints built into the design by the kind of interaction parameter specified. Perhaps the easiest way to appreciate these constraints is to observe that the birth date of a cohort is given identically by date of the present period minus the present age of the cohort. This constraint can be overcome by choosing not to estimate all of the cohort parameters. We accomplish this task by a side condition setting cohort effect parameters equal to zero for cohorts born from 1897 to 1904 and those born from 1934 to 1943.[1] With these side conditions estimation can proceed by least squares and the antilog of the parameters estimated for Equation (3) can be taken as estimates of the coefficient of Equation (2).

Note that this estimation technique does not permit estimation of preschool ability or the rate of return on human capital directly. It only permits an investigation of the proportionate deviation of these values from the mean taken over all of the data.

For education and age effects, however, it is possible to transform these "effect parameters" back into rates of return to capital for all but the first education or age category as follows: if γ_k is the estimate of T for the k^{th} education category then:

$$\frac{\gamma_{k+1}}{\gamma_k} = \frac{\frac{T_{k+1}}{\overline{T}}}{\frac{T_k}{\overline{T}}} = \frac{T_{k+1}}{T_k} = \frac{\prod\limits_{i=1}^{k+1} (1 + r'_i)}{\prod\limits_{i+1}^{k} (1 + r'_i)} + (1 + r'_{k+1}) \quad .$$

Thus, it is possible to find one's way back to the coefficients of Equation (1) for r' and ρ while for R and A one must make do with measures of proportionate change.

If one believes the model as presented, then the relative values for period parameters indicate the rise and fall over time in the rate of return on human capital. If the model is estimated separately for blacks and whites, a comparison of the range of these parameters for the two groups should answer the question: Is the rate of return to black human capital more sensitive to economic conditions than is white human capital?

An investigation of age effects should reveal the relative effect of experience versus depreciation of human capital. A comparison of the pattern of the parameters between the races should yield information on the relative experience-value of jobs available to blacks and whites.

An inspection of education effects is also of interest. Does the rate of return vary with level? Perhaps not. Thurow computed elasticity separately for elementary school, high school, and college training and found the elasticities increasing roughly proportionately to the added years of education (Thurow, 1969). This finding would be consistent with a constant rate of return for an educational unit. If the rate of return is not constant, a comparison between the races should yield information on the differing pattern of rates of return to education. For example, is completing high school of less value to blacks than whites? If rates within races seem constant, then our model can be simplified to permit estimation of the two rates themselves and we can answer the question: Is there less payoff in education for blacks than for whites?

According to the model as described above, a comparison of cohort parameters should reveal any effects of the diffusion of preschool training via kindergarten, nursery school, etc. as well as any trends in innate ability accruing from improved nutrition, medical service, and the like. We shall find this an unsatisfactory explanation of the observed results.

Finally, we can investigate the adequacy of the model itself by comparing the variance between age-period-education groups in mean log income with the variance explained by our model.

If the model is not an adequate summary of the data, its only convenient elaboration consists in interacting education terms with age, period, and cohort terms. The latter interaction might be especially useful if there have been temporal trends in the quality of education. Age-education interactions might be useful if there is sufficient age segregation of jobs to create separate human capital markets by age. Beyond these elaborations, inadequacies in the model can only be explored by investigating residuals.

THE DATA USED IN THIS PAPER

Estimating the model just described requires construction of tables for males by age, education, income, and race for a series of cross-sections. Because the model investigates (among other things) cohort changes in income it requires that the surveys be evenly spaced in time, that a comparably fine age detail be available, and that there be sufficient surveys to yield a reasonable number of periods for most cohorts.

To satisfy these conditions within the collection of data available at the Wisconsin Center for Demography I chose to use the following files

1. For 1960 - The 1/1000 sample of the 1960 Census.
2. For 1962 - The Occupational Change in a Generation Survey (OCG).
3. For 1966 - The 1966 Survey of Economic Opportunity (SEO).
4. For 1968 - An extract for persons of the March CPS tape provided by the Bureau of the Census to the Bureau of Labor Statistics.
5. For 1970 - A similar extract for the March 1970 CPS.

With the exception of the missing year 1964, this collection of surveys approximates a biennial sampling of the population. Each survey is large enough to support two-year age categories. Consequently, two-year birth cohorts can be followed with each being observed at five points in time. (I ordered tabulations from the March 1964 CPS from the Bureau of the Census but they did not arrive in time for inclusion in this analysis.)

On the surface, this set of surveys seem reasonably homogeneous. All sources derive from Bureau of the Census data collecting activities. The interviews were processed in somewhat similar ways. A fairly consistent set of coding procedures was used.

There are, however, some rather striking sources for "methods variance" among them. The method of enumeration varies. The sampling schemes are quite different. The same information is elicited by rather different questions. These differences are known to yield rather different distributions for basic variables.

At an even more operational level, consider the difficulties encountered with the variable income. The model requires a measure of earnings rather than total income. But the Occupational Change in a Generation Survey provides only total income, and that variable is coded in idiosyncratic categories ending in $15,000 and above. From the 1960 Census sample, an earnings variable can be constructed by adding together wages and self-employment income but errors are introduced in that procedure by the unique coding scheme of that data set. There are $10.00 categories up to $10,000, where the category bounds shift to $1,000. CPS and SEO files, however, code earnings directly in single dollars with an open-ended upper category of $50,000.

Thus, despite the fact that each of the separate files is in its own right an important social science resource which has yielded -- and will continue to yield -- valuable returns to our knowledge of society, the files do <u>not</u> automatically aggregate to produce reliable indicators of social change. That sort of aggregation is possible only by giving detailed attention to replication or by taking advantage of a facility such as the CPS whose basic design was created to produce social indicators (Duncan, 1969).

THE FINDINGS

Parameters of Equation (2) were estimated by the method discussed above for all nonfarm males between the ages of 25 and 64 in the periods 1960, 1962, 1966, 1968, and 1970. Sample weighting of cases was used as provided in each survey separately. Income was converted to 1960 dollars.

The correlation over individuals for our model is .336 for whites and .304 for blacks, while the value of the square root of the correlation ratio estimated from age, period, education cell means is .366 for whites and .399 for blacks.

Education

Table 1 presents parameter estimates for educational effects for whites and blacks. Columns 2 and 4 present parameters in the form of Equation (2). The numbers represent the proportionate deviation from the mean income attributable to being in the specific educational category. Columns 3 and 5 represent the numbers transformed into rates of return for the additional educational unit. Recall that these rates represent the return to human capital through investment in the additional unit of education compared to the return through the investment in work experience and that quantity multiplied by an ability selection factor.

Table 1 about here.

Although effect parameters rise roughly linearly with increased education, there are notable rises and falls in the rate of return to an additional unit of education. In general, rates of return are highest for the completion of one of the traditional breaking points in the educational system, i.e., 8, 12, and 16 years. For both races the rates are also high for the 5-7 years level. Elsewhere I have argued that it may be reasonable to regard accomplishing more than four years of schooling as an indicator of functional literacy. The high returns for the 5-7 level lend further credence to that notion (Winsborough and Dickinson, 1971).

In interpreting a finding of this kind -- higher returns to the completion of a traditional break point -- it is common to discuss the importance of a credential, such as a diploma, in our increasingly bureaucratized society. The details of our model suggest an additional plausible explanation. Recall that the relative rate of return in this

Table 1. Parameters for Education by Race

	Whites		Blacks	
Educational Level	Effect Parameters	Rates of Return for Additional Unit	Effect Parameters	Rates of Return for Additional Unit
0-4	.2376	-	.5422	-
5-7	.5012	1.1094	.8068	.4880
8	.7359	.4683	.9848	.2206
9-11	.9333	.2682	1.0628	.0792
12	1.1783	.2625	1.4220	.3380
13-15	1.2980	.1016	1.5585	.0960
16	1.7000	.3097	2.0257	.2998
17+	1.6251	-.0441	3.0701	.5156

model is multiplied by an ability selection factor. Perhaps these breaking points are important ability selection points. Some credibility can be accorded this argument by observing that most colleges and many high schools (and perhaps grammar schools in the past) have criteria for graduation which are more stringent than are those for continuance. Frequently a higher grade point average is required for graduation and certain kinds of courses must have been successfully completed by graduation time.

It could be argued that the foregoing over-interprets the rises and falls in the rates of return. A linear pattern appears to fit the effect parameters rather well. Perhaps the rates of return are simply deviations of the inter-point slopes from the total regression, i.e., from a constant rate of return. We find, however, that using a constant rate of return reduces the explained proportion of the variance between cell means by several points. This decrement is certainly statistically significant but more importantly, in my judgement, it is large enough to warrant substantive interpretation. If one wished to simplify the model, however, it seems clear that a constant, rather than a declining, rate of return is appropriate.

What are the differences between the races in these parameters? First, effect parameters show a greater range for blacks than for whites. Second, differences seem most dramatic where the fewest people are affected (and hence the parameters are the most unstable). The return for 5-7 years is greater for whites and the return for 17+ is greater for blacks. The exception to this rule is the markedly lower rate for blacks in the 9-11 category. Even though blacks show a somewhat higher return for the completion of high school, it is insufficient to compensate for the markedly lower parameter at the 9-11 level. The return for moving from grade 8 through 12 is for whites .60 while for blacks it is .44. Perhaps white employers have a tendency to take more seriously a white applicant's claim to "some high school" than they would a black applicant's claim.

Age

Table 2 presents parameters for age. Effect parameters have the expected general form. They rise to a peak around age 40 and then decline as the effects of "depreciation" overwhelm gains through experience. A plot of rates of return by age suggest that a roughly linear decline is appropriate. It is interesting to note that the peak value for effect parameters occurs at roughly the same age for both races but the decline for blacks appears to begin about five years earlier than it does for whites. Again, we find a greater range of parameters for blacks than for whites.

Table 2 about here.

Period

Table 3 presents effect parameters for periods. Recall that these parameters represent proportionate deviation in the rate of return to human capital and not the period-specific rates themselves. Thus, our model does not permit us to compare the rates themselves between the races but only permits us to investigate their comparative responsiveness to changing economic conditions.

Table 3 about here.

Table 2. Parameters for Age by Race

	White		Black	
Age	Effect Parameters	Rates of Return	Effect Parameters	Rates of Return
25-26	.6592	-	.7454	-
27-28	.8564	.2992	1.0242	.3740
29-30	.9529	.1127	.9641	-.0587
31-32	1.0635	.1161	1.0846	.1250
33-34	1.0895	.0244	1.3450	.2401
35-36	1.2126	.1130	1.1202	-.1671
37-38	1.2726	.0495	1.3260	.1837
39-40	1.3157	.0339	1.6727	.2615
41-42	1.2839	-.0242	1.7000	.0163
43-44	1.3050	.0164	1.5289	-.1006
45-46	1.2887	-.0125	1.2449	-.1858
47-48	1.2432	-.0353	1.2100	-.0280
49-50	1.2155	-.0223	1.2100	.000
51-52	1.1787	-.0303	.9648	-.2026
53-54	1.0813	-.0826	.8021	-.1686
55-56	.9891	-.0853	.7493	-.0658
57-58	.7868	-.2045	.6072	-.1896
59-60	.7018	-.1080	.4547	-.2512
61-62	.4880	-.3046	.3416	-.2487
63-64	.2917	-.4018	.2478	-.2746

Table 3. Parameters for Periods by Race

Period	White	Black
1960	.8245	.6404
1962	1.2143	1.2599
1966	.9589	.9898
1968	1.0512	1.1851
1970	1.0374	1.1130

In general, we observe the rising rates of return which would be expected over the period in question. The parameters for 1962 are the marked exception to the general pattern. Of course, a positive deviation from any trend would be expected for 1962 since the dependent variable in that period is total income rather than earnings. Because of this confounding of "methods" with "real" effects it is difficult to say whether 1962 represents a real departure from the trend. Between the races we again find a greater responsiveness in parameters for blacks than for whites. This finding, of course, is consistent with the last hired-first fired effect.

Cohorts

Table 4 presents effect parameters for cohorts. In the presentation of our model we asserted that it was the preschool ability factor which should vary by cohorts. We expected changes in this parameter to reveal secular trends in kindergarten attendance, etc., as well as changes in ability due to improved level of living, improved medical services to children, better nutrition and the like. The pattern of these parameters, however, makes this kind of interpretation unreasonable. First, the parameters are higher for older black cohorts and perhaps for white cohorts also. They decline in the middle and finally rise to their highest levels for the youngest cohorts. But those cohorts are ones whose preschool experience was during the Great Depression, hardly a time of improved level of living.

Table 4 about here.

Table 4. Parameters for Cohorts by Race

Birth Year	White	Black
1904-05	.8615	1.3256
1906-07	.9919	1.1775
1908-09	.9398	.9203
1910-11	.9219	1.2326
1912-13	.9255	.9497
1914-15	.9509	.9624
1916-17	.8852	.9444
1918-19	.9384	.8285
1920-21	.9461	.7441
1922-23	.9399	.8429
1924-25	1.0475	.7516
1926-27	.9732	.7834
1928-29	1.0009	.9573
1930-31	1.0377	1.0310
1932-33	1.0929	1.0859

What kind of alternative explanation can we produce for these results? Perhaps we have a cohort effect in the original meaning of that phrase. Perhaps variation in these parameters is negatively related to the relative size of the birth cohort. Keyfitz has recently presented cogent arguments to suggest that the economic advantage of being born into a small cohort and the disadvantage of being born into a relatively large one should be marked and persist throughout a lifetime (Keyfitz, 1972).

The advantage of being in a small cohort is presumed to accrue to a person through his entering various levels of education at times when the schools are not crowded, entering the labor market with fewer competitors, and achieving promotions more rapidly because of a scarcity of appropriately aged alternatives. Probably the best index of this condition would be a sequence of variables for each cohort showing, at various ages, the proportion of the population which is older. Rather than construct such a series of variables I simply aggregated white births and white population to two year groups from Coale and Zelnick and calculated the birth rate which "produced" each cohort (Coale and Zelnick, 1963:21-23). The correlation between the rates and the white cohort parameters is .78.

It is probably not worthwhile to perform the same computation for blacks since similarly reliable estimates for births and population are not available for many of the periods. An inspection of the available data, however, suggests a good correspondence.

A Summary of Black-White Differences

For every set of effect parameters we find a greater range of effects for blacks than for whites. For education and experience, perhaps there is a greater requirement for blacks than for whites to show some outward and visible sign of their human capital. For period and cohort effects, a last hired-first fired effect may be the most reasonable explanation wherein the behavioral motivation is simply racial prejudice.

An alternative explanation which has the virtue of explaining all four findings at once comes to mind. Perhaps primarily white employers have greater difficulty estimating the level of human capital for blacks than for whites. One might imagine the cultural and interpersonal separation of the races would make for great difficulty in assaying more subtle cues to ability level between races in the United States. Thus, one might expect that employers would lean more heavily on well-established criteria such as education and experience for blacks. At the same time, one response to economic recession or an over-supply of labor in a given cohort might be to reduce average uncertainty about the fit of ability to a given job. Such a response would account for the greater responsiveness of period and cohort for blacks since the employer's confidence interval about his estimate of a black's ability is presumed larger than it is for a white.

Investigation of Residuals

Overall the model predicts cell means rather well for whites and less well for blacks. Is there some pattern among the misestimations which would be helpful in revising the model? Several points of interest arise on the inspection of residuals.

First, there is a clear pattern of misestimation for 1962. For both races, income is underestimated at the older ages. Income is also overestimated for higher educational levels. It seems likely that this pattern is another effect of using total income rather than earnings for that period.

Second, for whites in 1960 there is exactly the reverse pattern of misestimation. The model overestimates at older ages and underestimates for higher educational levels. I think this pattern may derive from problems in measuring other nonwage and salary earnings in 1960. Particularly, I think the method of adding wage and salary income to self-employed income necessary with the 1960 1/1000 tape may lead to discounting of the importance of the self-employed income component of earnings.

Third, there is a rather clear pattern of education-cohort interaction for both races. For more recent cohorts income is overestimated at the higher levels of education and underestimated at the lower levels. The pattern is such as to suggest a decline over cohorts in the average rate of return to a unit of schooling. One interpretation of this finding might be that the supply of well-educated people is rising faster than the demand for them. There is another interpretation which I think better fits history. The cohorts we are inspecting were educated during the drive for "universal" grade school and then high school completion. That process has surely resulted in a change in the ability selection factor which, in our model, modifies the rates of return to education itself. Thus, it seems to me that the cohort decline in returns to the completion of higher educational levels is likely to result from declining ability section over time.

Finally, there is no pattern of errors for blacks which is not extant for whites. Therefore, I find no simple explanation for the somewhat greater predictive power of the model for whites.

<u>How CPS Files Would Have Been More Satisfactory</u>

In several instances it has been necessary to explain findings by recourse to the peculiarities of one of the surveys. For 1962, period effects were overestimated and inspection of residuals revealed a pattern of deviation of actual and predicted income by education and age. I explained these outcomes as a result of the use of total income in that period. In 1960, I found a pattern of education and age deviations which suggests these data may have discounted the nonwage and salary component of earnings in that period. Perhaps that explanation would also account for the rather low period parameter for 1960.

How have these obvious problems affected the estimation of other parameters? It is difficult to say with a model as complex as this one, but there has surely been an effect on the estimated age and education parameters as well as on the period ones.

As one considers those methods effects which have <u>not</u> made themselves obvious in the analysis, uncertainty increases even more rapidly. For example, the surveys are likely to have rather different response rates for blacks, and the pattern of selection bias in the returns among the surveys is probably not negligible. Does this fact account for the less satisfactory fit of our model for blacks? Perhaps so, but it would be a shame to allow the possibility of an additional factor in the model for blacks to go unexplored if that explanation is wrong.

SUMMARY OF THE ARGUMENT

The point of this paper has been to make an argument by example. The argument is that the archive of CPS files represents an important resource for the construction of social indicator models. The advantage of the CPS files consists in their inherent attention to comparability over time and the fact that numerous "replications" are available at comparatively low cost.

I have tried to illustrate the importance of attention to comparability in the details of data collection by counter example; that is, by showing how a series of "distinguished" cross-sectional surveys which appear marginally heterogeneous as to method yield marked ambiguities in interpretation of the analysis performed upon them by virtue of that heterogeneity.

I have tried to illustrate the advantage of numerous replications by choosing to use a model which captures inter-cohort change. Short of having genuine panel data, the capacity to deal in cohort change seems to me our primary hope for the estimation of dynamic social models. This capacity, however, depends on having several cross sections and its potentialities increase rapidly as more replications are added. The reader should note that the force of this argument depends neither on his agreement with the specific form of the cohort model here used nor on the adequacy of the estimation methods I have used. Rather, the point is that numerous replications permit specification and estimation of <u>some</u> cohort model.

Illustration of the relative cost advantage of using CPS files is perhaps the most ambiguous argument made here. My work was performed at an average cost of perhaps $10,000 while OCG II, reported on elsewhere in this volume, will cost over a million dollars for a simple replication. The comparison is, of course, unfair because the kinds of information to be collected in OCG II are not to be had in any other way but by replication and the value of having that information is high indeed. However, the information potentially available in the CPS also has a high value.

There is also another way in which the above cost comparison is unfair. Although the average cost of my work was relatively small, it was accomplished in the context of a research facility within the Center for Demography at the University (supported by a "Centers Grant" from NICHD) which is designed to minimize the time and money cost of just this kind of work. Without these facilities, the average costs would have been increased many-fold. To routinely and economically investigate social indicator models using CPS data would require design of a facility even more specialized than the one available to us at Wisconsin. Because I believe investment in such a facility would yield high returns to our knowledge of social change, I will conclude this paper with some observations on what I think would be required.

OBSERVATION ON A FACILITY TO PRODUCE INDICATORS FROM CPS FILES

Let me begin this section by recounting briefly the data processing and computation strategy employed in estimating parameters of the model previously described. As we began -- about the beginning of April -- it was unclear how many of the surveys we would actually be able to use. The 1968 and 1970 CPS person files were due to arrive shortly from the Rand

Corporation. We had ordered from the Census Bureau material from the 1964 CPS. We had never used our copy of the 1966 SEO. I chose, therefore, a computing strategy which allowed us to begin with those data at hand and "add on" additional periods as they become available. Therefore, for each of the surveys on hand -- the 1960 1/1000 Sample, the OCG, and the 1966 SEO -- we began selecting cases in the universe and constructing tables of age by education by race. We accumulated the number of observations in each cell, the sum of log income, and the sum of log income squared. All of these operations used the weighting scheme of the specific survey. As each of these tables were made they were stored on disk in a common format. At the same time, we designed a program to produce the required X'X and X'Y matrices from the tables. This program has the capacity to expand the number of periods fairly freely. Finally, we wrote a program to do the estimation itself. This program selects rows and columns from the previously constructed matrices. (It thereby permits some modification of the basic model and also will accommodate new surveys.) It then solves the equations under the constraint required and takes care of transforming these estimates into usable form.

In general, I think this computing routine is similar to one which would be chosen by a facility designed to estimate models from CPS files -- producing a new set of estimates with each new relevant survey. If I am right, two points in our experience deserve special attention.

First, the expensive part of the job was the data processing task. Of course, that task was less amenable to a generalized procedure in our work than would be a sequence of CPS files because most of our tapes were in different formats. On the other hand, we were dealing with only five files whereas between March, 1960, and March, 1970, there are extant 120 files of CPS surveys. Managing and processing that volume of data would be quite expensive. It would require a heavy personnel investment to keep track of the abundance of riches and very large computer bills would be generated as well.

Indeed, designing a data management system to retrieve information from the sequence of CPS files oriented to experiments in social modeling is a rather complex task. Such a system should keep track of rotation groups and primary sampling units (to facilitate estimating standard errors of parameters) as well as periods. A master codebook is required to indicate in which periods and for which rotation group a given set of questions were asked. Ideally, this codebook should itself be a machine readable file -- one which could be efficiently and accurately updated -- and should contain pointers to the data file. Overall, then, our experiment suggests that data management and data processing are difficult nonsubstantive tasks to be faced in designing a facility to experiment with models of social indicators using CPS files.

A second kind of computing problem which deserves comment arose in the course of our experiment. In the course of parameter estimation we encountered fairly difficult numerical analysis problems. Specifically, our X'X matrix contained numbers of widely varying magnitude. This fact, which seems endemic in work with files of the kind under discussion, can lead to a considerable problem with round off and accumulation errors. The estimation procedure for our model, for example, is very sensitive to an indexing error which will produce a matrix whose determinate is zero. We found that round off errors can interact with this kind of program error in fiendish ways -- ways which lead to indications of the difficulty designed to drive one out of his mind. At one point, we found ourselves dealing with a matrix whose determinate should have been positive but was in fact negative. To explore the difficulty, we computed Eigen values for the matrix. Some of them were negative (none of them should have been), and we thought we had a clue to our difficulty. However, since we had computed a negative determinate, we expected an odd number negative Eigen value. In fact, we found an even number.

We believe we have uncovered all the "bugs" in our programming and dealt with the round off and accumulation errors in a fashion comparable with the state of the art. Our experiment suggests, nonetheless, that a facility designed to do such work routinely should include skilled numerical analysts who are well versed in the adequacies and difficulty of the locally available package of programs for matrix algebraic computation.

The burden of these observations is that, presuming a decision were made to release CPS files for public use, the next problem would rapidly become that of designing and supporting a facility to make them usable. Our experience is that such a facility will require a rather high level of skill in both data management and numerical analysis.

REFERENCES

Coale, Ansley and Melvin Zelnick.

1963 New Estimates of Fertility and Population in the United States. Princeton: Princeton University Press.

Duncan, O. D.

1969 Toward Social Reporting: Next Steps. New York: Russell Sage Foundation.

Keyfitz, Nathan.

1972 Oscillations in a demographic-economic model. Paper presented at the Conference on Population Dynamics (June), sponsored by the Mathematics Research Center, University of Wisconsin.

Thurow, Lester.

1968 The occupational distribution of returns to education and experience for whites and Negroes. In Federal Programs for the Development of Human Resources Volume I. Washington, D. C.: Subcommittee on Economic Progress of the Joint Economic Committee, Congress of the United States.

1969 Poverty and Discrimination. Washington, D. C.: Studies in Social Economics, The Brookings Institute.

Winsborough, H. H. and Peter Dickinson.

1971 Components of Negro-white income differences. Proceedings of the Social Statistics Section, American Sociological Association:6-8.

FOOTNOTE

1. For a concise discussion of estimable functions, see Henry Scheffe, <u>The Analysis of Variance</u>, New York, John Wiley, 1959, pp. 13-19. For a more detailed discussion of the problem of estimability of cohort effects specifically, see Karen Mason <u>et al.</u>, "Some Methodological Issues in Cohort Analysis of Archival Data," <u>American Sociological Review</u>, forthcoming.