University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

Robert M. Hauser

ON "STRATIFICATION IN A DUAL ECONOMY"

DP #592-79

On "Stratification in a Dual Economy"

Robert M. Hauser

The University of Wisconsin-Madison

January 1980

This research was supported by National Institute of Mental Health Grant No. MH-06275. NORC-GSS data were obtained from the Data and Program Library Service of the University of Wisconsin-Madison. Computations were carried out using facilities of the Center for Demography and Ecology, provided with core support from the National Institute of Child Health and Human Development (Grant No. HD-05876). I thank Shu-Ling Tsai for extensive and careful research assistance. E. M. Beck has kindly responded to my requests for additional information about his research. I thank Glen Cain, O. D. Duncan, Neil Fligstein, Arthur S. Goldberger, Charles Halaby, Susan Janssen, Arne Kalleberg, Robert Kaufman, Robert D. Mare, William Mason, Rachel Rosenfeld, Aage Sørensen, Seymour Spilerman, Bruce Stephenson, and Ross M. Stolzenberg for their helpful reactions to an earlier draft of this paper. I am responsible for any errors in this report.

ABSTRACT

This paper reanalyzes NORC General Social Survey data used in a much-cited paper by Beck, Horan and Tolbert, "On Stratification in a Dual Economy," (1978) which purports to show that earnings processes are heterogeneous across "core" and "peripheral" industrial sectors. The reanalysis shows that Beck et al. reached this incorrect conclusion because they included improperly classified nonearners in a logarithmic earnings function. Other theoretical and methodological defects of their analysis are also discussed.

Beck et al. (1978) contrast what they term "individualistic" and "structural" theories of the distribution of earnings in an analysis of data from the 1975 and 1976 General Social Surveys (GSS) of the National Opinion Research Center (NORC). The purposes and results of their study are aptly summarized in their conclusion (p. 717):

> Our analysis examines the existence and the importance of industrial sectors, as hypothesized by the dual economy literature, on the process of earnings determination. Since the status attainment and neoclassical income models rest on the assumption of labor market homogeneity, our efforts to test a theoretically-derived model of nonhomogeneity are of direct relevance for those research traditions. Using a distinction between core and periphery industrial sectors derived from Bluestone et al., (1973) we address our analysis to three basic questions: (1) Are there significant differences in the economic status and composition of the core and periphery labor forces? (2) Can sectoral differences in economic status be reduced to differences in the quality of their respective labor forces? And, (3) to what degree are the economic returns to sex, race, human capital, and occupation-labor force variables the same in both sectors?
>
> The analysis provides clear answers to all three questions. The core and periphery sectors <u>do</u> exhibit significant differences in both earnings levels and in labor force composition. The sectoral differentials in earnings <u>cannot</u> be explained away

> by differences in labor force quality. The relationships between earnings and human capital as well as occupation-labor force variables do differ significantly between core and periphery sectors. Whether considering the biologically-fixed attributes of race and sex, the human capital variables, or the occupation-labor force variables, we find that the real dollar returns on these worker characteristics are greater in core industries than in periphery industries.
>
> The importance of these findings for the neoclassical research tradition in stratification should be clear. In face of strong empirical evidence contrary to the implicit tenet of labor market homogeneity, parameter estimates which are conditional on this assumption can no longer be treated as appropriate bases for the construction of sociological theory or social policy. Such simplistic models lead to a serious misspecification and misrepresentation of the social processes underlying individual earnings determination. In contrast, the notion of economic sectors appears to hold substantial promise as a theoretical concept and as an exemplar for a research program aimed at identifying structural aspects of the socioeconomic order.

There is no reason to question the first two of Beck et al.'s three major findings; industrial differentials in labor force composition and remuneration are well known. Their analysis and its conclusions focus primarily upon the third finding, "the relationships between earnings and human capital as well as occupation-labor force variables do differ significantly between core and periphery sectors." For example, Beck et al.

(p. 706) say, "the present analysis will focus on the issue of fixed returns--the assumption that economic returns to worker characteristics are uniform." On this point their analysis is wrong. Beck et al. incorrectly classify a small number of GSS respondents as zero earners. These few observations dominate their regression analyses. As a result, Beck et al. "detect" several nonexistent interaction effects; they also fail to detect one important interaction effect. Their regression analyses are wrong in virtually every detail.

Beck et al. begin their analysis by imputing "individualistic" assumptions of perfect competition and labor market homogeneity to research in the "neo-classical" human capital and status attainment traditions. The economists can speak for themselves; for an enlightening statement, see Cain (1976). I think these theoretical assertions about status attainment research are patently inaccurate. For example, in the sociological works cited by Beck et al. (705) as "individualistic in orientation and execution"--Hauser and Featherman (1977), Sewell and Hauser (1975), and Duncan et al. (1972)--one need not look far to find analyses of heterogeneity in the attainment process across populations defined by such variables as race, sex, age, calendar year, city size, and migration status, whose sociological explanation is not "tied directly to the characteristics brought into the marketplace by the individual workers" (Beck et al., p. 705). Far from assuming fully competitive and homogeneous labor market processes, a major thrust of recent stratification research has been to demonstrate the variety of social and psychological mechanisms--extending over the life cycle and affecting major social groups in different ways-- which tend to complicate and to invalidate such narrow theoretical frameworks.

The representation of labor market sectors in earnings functions is a straightforward theoretical and empirical elaboration of previous status attainment models (Featherman and Hauser 1978: Ch. 9). In fact, the Duncan SEI (Reiss 1961) incorporates a much larger number of distinctions of socioeconomic standing by industry within occupation than does the prestige scale employed by Beck et al. However, in attempting to elucidate the effects on earnings of industrial sectors as defined by Beck et al., I do not want to imply my agreement with their concept that there are just two economic sectors, with their assignment of each industry to "core" or "periphery" sectors, nor with their use of industries, rather than firms, to designate labor market sectors. I suspect that important inter-firm and inter-industry differences in earnings are obscured in the analysis of Beck et al.

To avoid confusion, I use "sectoral interaction effects" to refer to differences in slopes in semi-logarithmic earnings functions, like those in the left-hand panel of Table 3 of Beck et al. (p. 713). I do not refer to "real dollar" differences in the effects of variables as interaction effects. In this respect I believe that I am following the intent of Beck et al. (see p. 712, paragraphs 3 and 4, and p. 714, paragraphs 3 and 4). However, they also take note of "real dollar" differences in effects at some points in the text (pp. 714, 715, 717). In the semi-logarithmic earnings function each unit change in a regressor leads to a proportionate change in earnings, so the effect of a regressor in dollars depends on the value of the regressor at which its effect is evaluated (and on the values of other variables in the function). In the dollar earnings function, each unit change in a regressor leads to the same change in dollar earnings, regardless of the value of the regressor (or of other variables in the function). In the population, so long as sectors differ in intercepts of the earnings function, constant non-zero slopes in the semi-logarithmic functions imply differences in real dollar effects, and constant non-zero dollar effects imply different slopes in the semi-logarithmic functions. If one accepts both forms of interaction as evidence of heterogeneity in the labor market process, it is not logically possible to reject the hypothesis of heterogeneity.

There are good reasons to choose the semi-logarithmic form of the earnings function, which is widely applied in economics and sociology. First, it has a convenient interpretation--that variables have proportionate effects on earnings. Under some circumstances, such effects can be interpreted directly as rates of return (Mincer 1974). Moreover, one may expect proportionate effects to be invariant with respect to global price and productivity changes--or even to changes in units of currency across societies, while effects in raw units of currency will not be invariant. Second, the semi-logarithmic form of the earnings function has desirable statistical properties relative to the linear (dollar) form. The residuals (deviations from the regression line) are typically less skewed and less heteroscedastic when the semi-log function is applied. There is also evidence that the semi-logarithmic function fits earnings data better than the linear (dollar) function or other power functions of earnings (Heckman and Polachek 1974).

I think it unfortunate that Beck et al. identify heterogeneity in labor market processes with heterogeneity in regression slopes, for that identity holds only where there is a strong theoretical justification for the functional form in which heterogeneity appears. For example, the statisticians Mosteller and Tukey (1977) invoke parsimony as a criterion in the choice of functional forms; a researcher may radically transform a dependent variable in order to obtain good fit with a simple function of explanatory variables. In the context of such a flexible statistical analysis, heterogeneity of slopes is but a methodological nuisance. Obversely, one might infer substantive heterogeneity from functions with homogeneous slopes.

In my view Beck et al. offer no substantial theoretical rationale to predict the occurrence or form of a specific statistical interaction in their regression analysis. Moreover, they provide no theoretical rationale for semi-log or linear forms of the earnings function that would support the inference of heterogeneity in the earnings process, regardless of empirical findings. Indeed, while Beck et al. (1978) use a semi-log function to argue for sectoral heterogeneity, the same authors elsewhere use a linear function to make the same argument (Tolbert et al. 1980).

Beck et al. (p. 711) report that the mean logarithm of earnings is 8.5993 in the core with a standard deviation of 2.1148, and the mean is 7.7706 in the periphery with a standard deviation of 2.5628. These numbers are not credible. For example, they imply that 1975 dollar earnings at the mean in the core are \$5,428 (= $e^{8.5993}$); earnings at the mean in the periphery are \$2,370. As a rough approximation, the logarithm of earnings is normally distributed; thus one would expect dollar earnings at the mean of the distribution of the log of earnings to fall near the median of the distribution of dollar earnings. In the cumulative codebook of the GSS the median current dollar earnings of respondents are roughly \$7,400 in the 1975 and 1976 GSS (NORC 1977:45); even in the core the figure reported by Beck et al. is about \$2,000 less than one would expect from the assumption of lognormality. That is not all. At one standard deviation above the mean, their figures imply annual earnings of $44,986 in the core and $30,742 in the periphery; the highest category in the NORC earnings reports is "$25,000 or more.". At one standard deviation below the mean they imply annual earnings of only $655 in the core and $183 in the periphery. Something is wrong.

What accounts for the anomalously low means and large standard deviations in the distributions of the log of earnings? One reason to take the logarithm of earnings is to reduce the positive skew of the distribution. In compressing the high end of the dollar metric, the logarithmic transformation also stretches out the low end of the distribution (Mosteller and Tukey 1977:Ch. 4). Thus, the anomalies might be explained by observations of no earnings or very low earnings. The log of zero is undefined, so it is a common practice to assign $1 (or some other nominal sum) to zero earners before taking logs; when $1 is assigned the log is 0 by construction. This rule of thumb may generate extreme outliers (and it is made yet more dangerous by real growth or inflation). There are a number of better practices: add a larger nominal value, say, $500 or $1,000, to each observation; use a functional form which gives special treatment to zeros (Tobin 1958, Sewell and Hauser 1975:Ch. 6, Goldberger 1964:253-255); ignore the outlying observations (Featherman and Hauser 1978:Ch. 5); or choose a less radical transformation of earnings (Jencks et al. 1979:283-284).

The GSS uses a closed-ended question to ascertain earnings in the calendar year before the survey date. Each respondent chooses a broad income category from a card offered by the interviewer. The lowest category on the card is "under $1,000." Even if one assumed a uniform earnings distribution from $0 to $1,000, the midpoint of the interval would be $500. If log(500) = 6.21 were the minimum of the distribution of the log of earnings, that could not account for the anomalous observations.

There are 99 assigned values of zero in the distributions of the log of earnings used by Beck et al.; Beck has acknowledged this in a private communication. Moreover, these assigned values of zero do not occur primarily (or even secondarily) among persons holding low-paying jobs in the year before the survey date. They occur among persons who did not work at all in the year before the survey date or who changed occupations between the end of that year and the survey date (usually in March or April). Not only do the values of zero on the log of earnings give the distribution extreme negative skew, but the outliers have no clear connection with the occurrence of low pay.

The documentation of the GSS suggests how Beck et al. made this error. By assigning zeros to the log of earnings in certain cases from the GSS, I have been able to approximate their results rather closely. When the observations with assigned zeros are removed from my analysis, the anomalies disappear. So do the major findings of Beck et al.

In the 1975 and 1976 GSS respondents were classified as in the labor force if they were employed full or part-time, with a job but not at work, or unemployed (NORC 1977: Q. 1, p. 13). These and other persons with work experience were asked about their present or past occupation: "What kind of work do you (did you) normally do? That is, what (is/was) your job called?" (NORC 1977: Q. 2A, p. 15). At Q. 35 respondents were asked (NORC 1977: pp. 44-45):

35. Did you earn <u>any</u> income from (OCCUPATION DESCRIBED IN Q. 2) in [1972/73/74/75/76]?

Yes . . . . (ASK A)
No . . . . [See <u>REMARKS</u>]

A. <u>IF YES</u>: In which of these groups did your earnings from (OCCUPATION IN Q. 2) for last year--[1972-1976]--fall?

That is, before taxes or other deductions. Just tell me the letter.

REMARKS: Card E contained responses for punches 01-12 only. Q. 35 responses are not in these data, but are contained in Q. 35-A.

Question 35 says that persons with no earnings from the occupation described in question 2 were not asked to report personal earnings in the previous year; their earnings were coded as "blanks" for "not applicable" in the GSS file. The limitation of the personal earnings question to earnings in the named occupation is reinforced by the interviewer instructions (NORC 1977:179):

1976

Note instruction: SEE Q. 2. IF ANY OCCUPATION RECORDED, ASK Q. 35. Read job title from Q. 2. Record R's income only from job described in Q. 2, not from any other jobs or sources. This question is concerned with OCCUPATION, rather than specific job or employer. EXAMPLE: If R's occupation was "waitress," you would ask for total income in 1975 as a waitress. . . even if she worked in several different places during the year.

1974-75

NOTE INSTRUCTIONS. Read job title from Q. 2. This question is concerned only with any income R. earned from occupation described in Q. 2, not any other income from other sources or income of other family members.

Beck et al. treated current members of the labor force who had no earnings in the named occupation in the previous year as if they were zero earners by replacing "blanks" in the dollar earnings variable with zeros in the logarithmic transformation of earnings. This procedure implies that persons who were not in the labor force at all in the previous year were treated as extreme low earners in that year. It implies that persons who changed occupations after the beginning of the survey year were treated as extreme low earners in the previous year. It implies that persons who changed occupations during the previous year had only part of that year's earnings covered by the GSS personal earnings question. To be sure, some persons may have been unemployed throughout the previous year and thus have been classified legitimately as in the labor force but without earnings in that year. I think such persons are small in number relative to those improperly classified as non-earners, and it would in any event be questionable to assign zeros to such persons on the log of earnings.

Unfortunately, the GSS data lack the information required to enumerate persons properly and improperly classified as non-earners. Given the broad and heterogeneous make-up of the sample selected by Beck et al., which included all women in the labor force at the survey date and did not impose an upper age limit, I think it probable that most non-earners were improperly classified. Of 1,695 GSS respondents in the labor force at the survey date, Beck et al. (p. 708) excluded 12 for whom industry was not reported. Of the remaining 1,683 persons in their sample, 99 were assigned "blanks" on personal earnings in the previous year--53 in the core and 46 in the periphery.

The treatment of the 99 zero earners in the labor force, who make up less than 6 percent of the sample used by Beck et al., has a critical effect on the results. For example, Table 1 shows the means and standard deviations of variables used in my reanalysis of the GSS data. LNY1 is my approximation of the dependent variable employed by Beck et al. That is, I adjusted the 1974 data to 1975 dollars, placed missing observations at the mode, assigned each person the log of the midpoint of his earnings interval, and used a Pareto approximation to estimate the mean of the upper, open-ended interval; the 99 observations with zero earnings were assigned zero values on LNY1. In the case of LNY2, the procedures were the same, except the 99 observations with zero earnings were ignored. The characteristics of LNY1 are very similar, though not identical to those of the earnings measure used by Beck et al. (p. 711). I have found other discrepancies in my reanalysis, but I think their effects are minor. The means and standard deviations of LNY2 are vastly different from those of LNY1 in both the core and the periphery. For example, in the core sector dollar earnings at the mean of LNY2 are $8,451, compared to $5,519 in the case of LNY1. At one standard deviation above the mean, earnings based on LNY2 are $20,517, but they are $45,207 when based on LNY1. At one standard deviation below the mean, earnings based on LNY2 are $3,481, but they are $674 when based on LNY1. Similar results are obtained in the periphery, where earnings at the mean of LNY2 are $4,969, and they are $2,463 at the mean of LNY1. At one standard deviation above the mean in the periphery, earnings based on LNY2 are $14,501, and those based on LNY1 are $31,793. At one standard deviation below the mean, earnings based on LNY2 are $1,703, and those based on LNY1 are $191.

In core and periphery the standard deviation of LNY2 is less than half as large as that of LNY1. In the total sample, earnings at the mean of LNY2 are $7,115; that figure is reassuringly close to the median current dollar earnings of $7,400. In the total sample, earnings at the mean of LNY1 are only $4,226. Thus, the presence of (inappropriate) zero observations accounts for anomalous characteristics of the earnings measure used by Beck et al.

TABLE 1: Means and Standard Deviations of Characteristics of Workers in the Civilian Labor Force by Industrial Sectors: 1975 and 1976 NORC General Social Surveys

Worker Characteristic		Total N=1683	Industrial Sector: Core N=1125	Industrial Sector: Periphery N=558
LN annual earnings	LNY1[1]	8.349 (2.295)	8.616 (2.103)	7,809 (2.558)
	LNY2[2]	8.870 (.982)	9.042 (0.887)	8.511 (1.071)
Sex (1=female)	X_1	0.403 (0.491)	0.370 (0.483)	0.471 (0.500)
Race (1=nonwhite)	X_2	0.102 (0.302)	0.092 (0.289)	0.122 (0.327)
Age of males	X_3[3]	39.39 (13.871)	39.632 (13.290)	38.810 (15.187)
Age of females	X_4[4]	38.507 (13.543)	37.911 (12.920)	39.449 (14.451)
Years of schooling	X_5[5]	(12.385) (2.988)	12.853 (2.975)	11.443 (2.788)
Highest degree				
High school or junior college	X_6	0.553 (0.497)	0.552 (0.498)	0.554 (0.498)
Bachelor's degree	X_7	0.131 (0.337)	0.158 (0.365)	0.075 (0.264)
Postgraduate degree	X_8	0.051 (0.220)	0.073 (0.260)	0.007 (0.084)
Occupational prestige	X_9[6]	39.298 (13.777)	42.255 (13.808)	33.358 (11.640)
Union member (1=yes)	X_{10}	0.244 (0.430)	0.297 (0.457)	0.138 (0.345)
Unemployed (1=yes)	X_{11}	0.071 (0.257)	0.073 (0.260)	0.068 (0.252)
Work stability (0=stable)	X_{12}[7]	0.336 (0.472)	0.332 (0.471)	0.344 (0.475)
Years parental schooling	X_{13}[8]	9.409 (3.895)	9.630 (3.930)	8.966 (3.787)
Father's prestige	X_{14}[9]	38.906 (12.216)	39.404 (12.478)	37.871 (11.599)

--Table 1 continues

Table 1--continued

[1]Cases with "zero" earnings are assigned LNY1=0, and missing earnings data are assigned the modal value, \$12,500 (N_T = 1683, N_C = 1125, N_P = 558).

[2]Missing earnings data are coded as \$12,500, but cases with "zero" earnings are excluded (N_T = 1584), N_C = 1072, N_P = 512).

[3]2 missing data for male ages are coded as 27 or 25 for core or periphery sectors, respectively (N_T = 1004, N_C = 709, N_P = 295).

[4]3 missing data for female ages are coded as 24 or 25 for core or periphery sectors, respectively (N_T = 679, N_C = 416, N_P = 263).

[5]4 missing data excluded (N_T = 1679, N_C = 1122, N_P = 557).

[6]4 missing data excluded (N_T = 1679, N_C = 1121, N_P = 558).

[7]7 missing data excluded (N_T = 1676, N_C = 1120, N_P = 556).

[8]Parental education is taken as the years of schooling for the respondent's father. If this information is not available, data for mother's education or modal values were substituted using an algorithm provided by Beck et al.

[9]200 missing data excluded (N_T = 1483, N_C = 1001, N_P = 482).

How do the zero observations affect the regression analysis? To answer this question, I carried out new regression analyses of earnings in the core and the periphery, first, using all observations with LNY1 as the dependent variable, and, second, using observations with non-zero earnings and LNY2 as the dependent variable. I did not reanalyze the earnings of the subsample of full-time employed white males (p. 716), nor did I reanalyze the linear probability function which Beck et al. used to estimate the incidence of poverty-level earnings (pp. 713, 716). I do not think a reanalysis of earnings in the male subgroup would add anything of substance to the present analysis; Beck has provided tabulations for that subgroup which show the same anomalies as the published tabulations for the full sample. The incidence of poverty-level earnings, like the overall distribution of the log of earnings, is affected substantially by the inappropriate classification of job changers and persons outside the labor force as non-earners or low earners. Also, statistical defects in the linear probability function are well documented, and these are likely to be important in analyses of a highly skewed variable like the incidence of poverty.

Table 1 reports the means and standard deviations of variables used in my reanalysis of the GSS data; the reader should compare it with Table 2 of Beck et al. (p. 711). For most variables the two analyses yield the same estimates. As noted in the table, I reproduced Beck et al.'s results for age by assigning modal values to a small number of missing observations; this procedure was not mentioned in the source. I was unable to resolve a minor discrepancy in the proportion of persons with a high school or junior college degree. In the core sector my estimate of mean occupational prestige is larger, and its standard deviation is smaller than that of Beck et al. The prestige distribution included four zero values, which are out of its legitimate range, and I dropped these from the analysis.[1] In the case of work stability I excluded nonresponses, which Beck et al. apparently did not do.

Father's occupational prestige deserves special mention. My estimates of its mean are larger and my estimates of its standard deviation are smaller than those of Beck et al. They randomly assigned one of three modes of the prestige distribution to substitute for missing data, so their result cannot be reproduced independently (p. 708).[2] My estimates were obtained by dropping the 200 cases where father's prestige was missing. Aside from the large number of missing observations, parental prestige had trivial negative effects and did not interact with economic sector in the earnings regressions (p. 715). For these reasons, I dropped it from the regression analysis.

Beyond the differences just noted, I have tried to repeat the analysis of Beck et al. for each version of the dependent variable (LNY1 and LNY2). In particular, I have not tried to improve the specification of the earnings function in other ways. I believe several defects in specification remain, so my results are of interest mainly because of their differences from those of Beck et al. For example, there is ample theoretical and empirical evidence that a linear term in age does not adequately represent the effect of work experience on earnings, especially in a sample with no upper age limit. Aside from lifetime work experience, it is also important to include more proximate measures of labor supply--weeks and hours worked--in the earnings function. No such measures were obtained for the previous year in the GSS, yet it is reasonable to expect that differences in labor supply are intricately connected with sectoral differentials in earnings, as they are with sex and race differentials in earnings. It is well established, also, that occupational socio-economic status is superior to occupational prestige in representing occupational differences in earnings. Further, heterogeneity in earnings functions across major social groups--like race, sex, age, and geographic location--has been a major focus of research. While Beck et al. do attempt to show sectoral heterogeneity among full-time white male workers, it is curious that they do not specify heterogeneity in the regressions by race or sex within labor market sectors. Indeed, were there not other reasons to question the validity of their regression analysis, it would be pertinent to ask whether differences in the race and sex composition of the core and periphery sectors may help to explain the observed heterogeneity across sectors in the function specified by Beck et al.

Table 2 shows the results of my efforts to reproduce the earnings regressions of Beck et al. using LNY1. Columns (a) and (b) of Table 2 show the earnings regressions in the core and periphery, respectively. Despite the differences between our analyses, my estimates are very similar to those reported by Beck et al. (p. 713). Column (c) reports estimated slopes in a pooled equation, where intercepts differ in the core and periphery, but common (homogeneous) slopes are estimated. This equation is not reported by Beck et al., but it is reassuring that the coefficients of male age, occupational prestige, union membership, and work stability do resemble the common slopes for those variables reported by Beck et al.

I also estimated a single pooled equation in which each sector has unique slopes and intercepts. Of course, the estimated slopes from that

TABLE 2: Regression analysis of earnings (LNY1) by industrial sector in the civilian labor force: LNY1=0 for "zero" earners, and missing earnings placed at the mode

Worker Characteristic		Sector-specific Core	Sector-specific Periphery	Pooled Equation	F Statistic for Sectoral Interaction	BHT Finding
		(a)	(b)	(c)	(d)	(e)
Intercept		7.438	5.183	6.671		
Sector (1=periphery)	X_0	--	--	-.3238** (.1180)		
Sex (1=female)	X_1	-.9006** (.3724)	-.1782 (.5626)	-.6064* (.3112)	1.259	sig
Race (1=nonwhite)	X_2	-.4696** (.2047)	.3228 (.3188)	-.2127 (.1737)	4.831**	sig
Age of males	X_3	.0192** (.0058)	.0183* (.0102)	.0192** (.0051)	.008	non-sig
Age of females	X_4	.0187** (.0075)	-.0015 (.0101)	.0108* (.0060)	2.761*	sig
Years of schooling	X_5	-.0606 (.0452)	.1538** (.0663)	.0230 (.0374)	6.625**	sig
Highest degree						
High school or junior college	X_6	.4513** (.2099)	.2810 (.3175)	.3830** (.1765)	.124	sig
Bachelor's degree	X_7	1.0106** (.3664)	-.7354 (.6232)	.4871 (.3161)	5.645**	sig
Postgraduate degree	X_8	1.4160** (.4753)	1.2456 (1.3147)	.9015** (.4222)	.001	sig
Occupational prestige	X_9	.0207** (.0055)	.0253** (.0099)	.0205** (.0048)	.065	non-sig

--continued

TABLE 2--continued

Worker Characteristic		Sector-specific: Core (a)	Sector-specific: Periphery (b)	Pooled Equation (c)	F Statistic for Sectoral Interaction (d)	BHT Finding (e)
Union member (1=yes)	X_{10}	.5310** (.1348)	.7440** (.2927)	.5636** (.1260)	.466	non-sig
Unemployed (1=yes)	X_{11}	-.6716** (.2389)	-2.0206** (.4216)	-1.0836** (.2122)	8.874**	sig
Work stability (1=stable)	X_{12}	.3855** (.1347)	.5749** (.2321)	.4619** (.1187)	.636	non-sig
Years parental schooling	X_{13}	-.0172 (.0173)	-.0816** (.0314)	-.0369** (.0155)	3.444*	sig
Standard error of estimate		1.939	2.337	2.094	2.079	
R^2		.160	.185	.174	.192	
N^b		1116	555	1672	1672	

*Significant at the 0.10 level

**Significant at the 0.05 level

[a]Approximate standard errors of coefficients in parentheses.

[b]These are minimum pairwise-present counts, so those for the sectors need not add to that for the total sample.

equation merely recapitulate the findings reported in columns (a) and (b); column (d) shows F-statistics for the heterogeneity in slope of each variable in the pooled equation with unique slopes. Also, the base of column (d) reports summary measures of fit of the equation with unique slopes.

While I think it fair and accurate to say that the results in columns (a), (b), and (c) reproduce the major features of the analysis by Beck et al., there are significant departures from their findings about sectoral interaction and about the fit of the regression equation. Note I am discussing my effort to <u>reproduce</u> their findings, <u>not</u> the effect of including misclassified non-earners in the regression analysis. Column (e) of Table 2 reports the findings of Beck et al. with respect to sectoral interaction in their semi-logarithmic earnings function. Beck et al. say that they reject the null hypothesis of homogeneity when interactions are significant at or beyond the $p = .10$ level (pp. 712-713), assuming the GSS is as efficient as a simple random sample (NORC 1977:157). My analysis follows this same generous rule, but in 3 of the 9 cases where Beck et al. report sectoral interactions (sex, high school, and postgraduate degree), my test statistics fall well <u>below</u> their critical values. These discrepancies also appear when I construct a t-statistic from the point estimates and standard errors of the slopes in equations (a) and (b); I have been unable to resolve them.[3] Incidentally, Beck et al. (p. 713) report $R^2 = .2962$ in an equation containing only the significant sectoral interaction effects, while I find $R^2 = .192$ in an equation containing all sectoral interaction effects. Beck has informed me that $R^2 = .2008$ in his analysis; the larger figure was obtained by correlating observed dollar earnings with the antilog of predicted log earnings.

Table 3 shows a regression analysis of the log of GSS earnings from

TABLE 3: Regression analysis of earnings (LNY2) by industrial sector in the civilian labor force: data for "zero" earners excluded and missing earnings placed at the mode

Worker Characteristic		Sector-specific Core	Sector-specific Periphery	Pooled Equation	F Statistic for Sectoral Interaction	BHT Finding
		(a)	(b)	(c)	(d)	(e)
Intercept		7.844	7.316	7.716		
Sector (1=periphery)	X_0			-.1889**		
				(.0453)		
Sex (1=female)	X_1	-.5241**	-.6847**	-.5902**	.363	sig
		(.1436)[a]	(.2155)	(.1195)		
Race (1=nonwhite)	X_2	-.0037	.0786	.0110	.332	sig
		(.0790)	(.1221)	(.0667)		
Age of males	X_3	.0099**	.0071*	.0094**	.672	non-sig
		(.0022)	(.0039)	(.0020)		
Age of females	X_4	.0067**	.0061	.0069**	.079	sig
		(.0029)	(.0039)	(.0023)		
Years of schooling	X_5	.0028	.0435	.0157	.010	sig
		(.0174)	(.0254)	(.0144)		
Highest degree						
High school or junior college	X_6	.2708**	.1989	.2302**	.128	sig
		(.0810)	(.1216)	(.0678)		
Bachelor's degree	X_7	.4511**	.1276	.3344**	.038	sig
		(.1413)	(.2387)	(.1214)		
Postgraduate degree	X_8	.6490**	1.0978**	.5139**	2.281	sig
		(.1833)	(.5035)	(.1621)		
Occupational prestige	X_9	.0113**	.0247**	.0158**	11.503**	non-sig
		(.0021)	(.0038)	(.0019)		

--continued

TABLE 3—continued

Worker Characteristic		Sector-specific Core	Sector-specific Periphery	Pooled Equation	F Statistic for Sectoral Interaction	BHT Finding
		(a)	(b)	(c)	(d)	(e)
Union member (1=yes)	X_{10}	.3292** (.0520)	.3600** (.1121)	.3169** (.0484)	.014	non-sig
Unemployed (1=yes)	X_{11}	-.2208** (.0921)	-.1009 (.1615)	-.1588* (.0815)	.608	sig
Work stability (1=stable)	X_{12}	.2909** (.0519)	.3170** (.0889)	.2915** (.0456)	.082	non-sig
Years parental schooling	X_{13}	-.0010 (.0067)	-.0484** (.0120)	-.0150** (.0059)	11.653**	sig
Standard error of estimate		.731	.858	.781	.776	
R^2		.329	.375	.374	.386	
N^b		1067	510	1577	1577	

*Significant at the 0.10 level

**Significant at the 0.05 level

[a] Approximate standard errors of coefficients in parentheses.

[b] These are minimum pairwise-present counts, so those for the sectors need not add to that for the total sample.

which the 99 observations assigned zero values have been dropped. In every other respect the analysis is unchanged, but the results differ dramatically from those reported in Table 2 and from those reported by Beck et al. First, the fit is vastly improved. As shown by the standard errors of estimate and coefficients of determination, the predictive value of the simple earnings function is substantially greater in the core, in the periphery, and in the total sample. For example, the pooled equation with interaction effects accounts for 19.2 percent of the variance in LNY1 and for 38.6 percent of the variance in LNY2; the unexplained variance of earnings is more than seven times larger in the case of LNY1 than in that of LNY2 ($2.079^2/.776^2 = 7.2$).

Second, when the outliers are dropped, there remain only two statistically significant interaction effects. The (negative) effect of parental schooling is significantly larger in the periphery than in the core. While this negative effect is statistically significant in the periphery and in the total sample, the sign of the coefficient makes no theoretical sense. I am inclined to regard that finding as evidence of measurement error--possibly the treatment of missing data--in parental schooling or as evidence of misspecification elsewhere in the earnings function; Beck et al. also have little to say about a parallel finding in their analysis.

More interestingly, the estimated effect of occupational prestige is more than twice as large in the periphery as in the core. This is the only interaction effect in Table 3 which may be worth interpreting, and it was not detected in the analysis by Beck et al. (pp. 713, 715) nor in that of Table 2. Both in the core and in the periphery the effects of prestige on earnings are substantial. Earnings increase by 1.1 percent for each unit of prestige (on the NORC scale) in the core, and they increase by 2.5 percent for each unit of prestige in the periphery. These different returns may be a result of nonlinearity rather than of heterogeneity, for jobs in the core and in the periphery cover different ranges of the prestige scale. Recall that mean prestige is 42.3 in the core and 33.4 in the periphery; the two means are nearly a standard deviation apart. I have not investigated this possibility. In the theoretical context developed by Beck et al., I think it may be difficult to explain why the effects of prestige are twice as large in the periphery as in the core. For example, Beck et al. write (p. 707):

> Theories of dual economy suggest that these sectoral differences have important implications for the opportunity structures and experiences faced by individual workers. In the core sector, workers move within job structures characterized by differentiated task and wage schedules with often well-defined career patterns, i.e., internal labor markets (Doeringer and Piore, 1971; Spilerman, 1977). Formal education is widely used to mediate individual access to job ladders, and workers' wages "are largely determined by their respective access to different

job clusters, by the relatively rigid pattern of wages attached to the job structures through which they respectively move, and by the speed with which they pass through those structures" (Gordon, 1972:50). In the peripheral sector, occupational opportunity structure is more restricted (sic) with a consequent dampening of task and wage variations. Gordon (1972:51) suggests that in this sector, "variation in individual hourly wages will depend very little on variations in individual 'capacities' like aptitude, reasoning and vocational skill."

Whatever the theoretical ramifications of the differential effects of prestige, I think it would be premature to draw global conclusions about heterogeneity in labor markets from the present finding.

Again, aside from the effects of occupational prestige and of parental schooling, the analysis in Table 3 detects no statistically significant interaction effects. This may be contrasted with 9 sectoral interactions reported by Beck et al. (of which I was able to confirm only 6). Some of the non-effects are worth mentioning. Beck et al. (p. 715) suggest that "earnings return to education in the core sector rests on the acquisition of a formal degree, whereas in the peripheral sector economic benefits are derived from additional years of schooling, not from increases in formal levels of certification." In the analysis of Table 3, the linear term in schooling is not significant in core or periphery, and the effects of levels of certification do not differ across sectors. Beck et al. (pp. 717-718) also write:

> We further contend that the sectoral model employed here constitutes an essential element in understanding the process of

> discrimination against minority groups in that it includes not only individual factors but also the organization of the economic structure. Specifically, despite the differential representation of females and nonwhites in the peripheral sector, there is no statistically significant evidence of earnings discrimination in that sector after controlling on the predetermining variables. In the core sector, however, there is evidence of significant adverse race and sex main effects on earnings even after controlling on human capital and occupational variables.

In the analysis of Table 3, women suffer a very large economic disadvantage, which is not significantly different in the core and in the periphery. Furthermore, with the other variables controlled, nonwhites have no significant advantage or disadvantage in the core or in the periphery, nor do the estimates of the effect of race differ significantly between the core and the periphery.

Not only are most of the revised estimates in Table 3 similar in the two economic sectors, but they are very different from those estimated in Table 2 and, similarly, from those estimated by Beck et al. I offer a few illustrations constructed by taking anti-logs of coefficients from the revised equation with common slopes (column (c) of Table 3) and from the unique sectoral equations estimated by Beck et al. Beck et al. estimate that women make 44 percent ($e^{-.8232}$ X 100) as much as men in the core and 87 percent as much in the periphery; in Table 3 I find that in both sectors women make 55 percent of the earnings of men. Beck et al. estimate that nonwhites make 63 percent as much as whites in the core and 38 percent <u>more</u> than whites in the periphery. In Table 3 I find no differences between the earnings of whites and nonwhites; this seemingly anomalous finding may be a result of the inclusion of both men and women in the analysis without a measure of labor supply. Beck et al. estimate that the currently unemployed made 50 percent as much in the previous year as the currently employed in the core, and the currently unemployed made 14 percent as much in the previous year as the currently employed in the periphery. In Table 3 I find that the currently unemployed made 79 percent as much in the previous year as the currently employed. My reading of these comparisons is that the estimates based on Table 3 are reasonably consistent with other social scientific findings, but those of Beck et al. are not.

Conclusion

Beck et al.'s analysis of sectoral differences in earnings, based upon data from the 1975 and 1976 NORC General Social Surveys, is fatally flawed. Because of an error in sample selection, many job changers and persons out of the labor force were treated as if they worked but had no earnings. This error was compounded by using a logarithmic transformation of earnings, which gives a great deal of weight to observations at or near the zero point of the transformed variable. Because of this error, their analyses of earnings (but not those of labor force composition) are wrong in important respects. In this reanalysis of the GSS data I have emphasized the effect of the error upon Beck et al.'s major conclusion, that the earnings process is heterogeneous across industrial sectors. When the error is corrected, most of the evidence of heterogeneity disappears.

To be sure, there remain differentials in earnings between core and periphery sectors. In the corrected equation of Table 3, where common slopes are estimated in the two industrial sectors, I estimate that workers in the periphery make 83 percent as much as workers in the core; the differential is $1,455 when evaluated at the mean of the earnings distribution in the core, and it is $1,033 when evaluated at the mean of the earnings distribution in the periphery. The resemblance between these figures and the estimates of Beck et al. (Table 4, p. 713) are purely coincidental. The equations used by Beck et al. to standardize earnings between sectors vastly overestimate the sectoral earnings differential in <u>relative</u> terms, but this error is largely compensated by erroneously low estimates of average earnings in both sectors.

Another respect in which the analysis of Beck et al. is roughly correct is the share of the total sectoral earnings differential which is explained by the variables included in their regression equation. In my revised estimates the difference in the log of earnings between core and periphery is .531; evaluated at the means in the two sectors, the difference is $3,482. Since the coefficient of sector in the pooled equation of Table 3 is -.1889 differences in social composition between sectors explain about two-thirds of the initial difference in earnings.

These points of agreement are conditional upon the specification of the earnings functions in Table 3. While I believe they do not share the flaws of the equations estimated by Beck et al., there remain several specification errors in those equations. In particular, I think that the absence of a proximate measure of labor supply--the amount of time at work in the previous year--explains part of the remaining sectoral difference in earnings.

I doubt that any methodological commentary can counter the impact of the broad and sweeping conclusion of Beck et al. (p. 717): "In face of strong empirical evidence contrary to the implicit tenet of labor market homogeneity, parameter estimates which are conditional on this assumption can no longer be treated as appropriate bases for the construction of sociological theory or social policy." It is thus most regrettable that their proposition finds so little empirical support.

Footnotes

1. According to Bruce Stephenson of NORC, these four cases were persons who reported jobs whose descriptions could not be coded.

2. I attempted to obtain the recoded file including this variable from Beck, but he reported that the file had been destroyed.

3. Beck has not responded to my inquiries about these discrepancies.

References

Beck, E. M., Patrick M. Horan, and Charles M. Tolbert II

1978 "Stratification in a Dual Economy: A Sectoral Model of Earnings Determination." American Sociological Review 43 (October):704-720.

Cain, Glen G.

1976 "The Challenge of Segmented Labor Market Theories to Orthodox Theory." Journal of Economic Literature 14 (December): 1215-1257.

Duncan, Otis Dudley, David L. Featherman, and Beverly Duncan

1972 Socioeconomic Background and Achievement. New York: Seminar.

Featherman, David L., and Robert M. Hauser

1978 Opportunity and Change. New York: Academic.

Goldberger, Arthur S.

1964 Econometric Theory. New York: Wiley.

Hauser, Robert M., and David L. Featherman.

1977 The Process of Stratification: Trends and Analyses. New York: Academic.

Heckman, James, and Solomon Polachek

1974 "Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship." Journal of the American Statistical Association 69 (June):350-354.

Jencks, Christopher, et al.

1979 Who Gets Ahead? The Determinants of Economic Success in America. New York: Basic.

Mincer, Jacob

1974 Schooling, Experience, and Earnings. New York: National Bureau of Economic Research and Columbia University Press. .

Mosteller, Frederick, and John W. Tukey

1977 Data Analysis and Regression: A Second Course in Statistics. Reading, Mass.: Addison-Wesley.

National Opinion Research Center

1977 National Data Program for the Social Sciences: Cumulative Codebook for the 1972-1977 General Social Surveys. Chicago: National Opinion Research Center.

Reiss, Albert J., Jr.

1961 Occupations and Social Status. New York: Free Press of Glencoe.

Sewell, William H., and Robert M. Hauser

1975 Education, Occupation and Earnings: Achievement in the Early Career. New York: Academic.

Tobin, James

1958 "Estimation of Relationships for Limited Dependent Variables." Econometrics 26 (January):24-36.

Tolbert, Charles, Patrick M. Horan, and E. M. Beck

1980 "The Structure of Economic Segmentation: A Dual Economy Approach." American Journal of Sociology 85 (forthcoming).