NBER WORKING PAPER SERIES

ESTIMATING THE DETERMINANTS OF EMPLOYEE PERFORMANCE

Charles Brown

Working Paper No. 353

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

May 1979

I have benefited greatly from the suggestions of seminar participants at these universities: Harvard, Maryland, Chicago, Washington (St. Louis), and Columbia. I am grateful to Bronwyn Hall for help with the maximum likelihood estimation package MAXLIK, to Nancy Lemrow for research assistance, and to the General Research Board and the Computer Science Center at the University of Maryland for supporting me and the computations, respectively. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper 353
May 1979

"Estimating the Determinants of Employee Performance"

ABSTRACT

Employers often wish to know whether the factors used in selecting employees do in fact allow them to choose the most qualified applicants. Because the performance of those not chosen is not observed, sample-selection bias is a likely problem in any attempt to "validate" employee-selection criteria. With minor modifications, the recently-developed techniques for dealing with sample-selection problems can be used in this context.

Using data on applicants for first-line supervisory positions and ratings of on-the-job performance of those hired, ordinary least squares estimates of the determinants of performance are compared with maximum-liklihood estimates which correct for selection bias. The correction for selection bias produces some appreciable improvements in some variables' coefficients, though the corrected estimates remain "insignificant" at conventional levels. Differences in the firm's stated and actual hiring criteria are also noted.

Charles Brown
Department of Economics
University of Maryland
College Park, MD 20742

(301) 454-4182

The relationship between job performance and individual attributes known by the employer at time of hiring is an important topic among economists. Selecting the best workforce from a pool of applicants is an essential function for a firm which hopes to survive competitive pressures. Evidence on how successfully firms perform this function is important for verifying the cost-minimizing behavior that most economic analyses assume.[1] On a more day-to-day level, Federal guidelines on non-discriminatory hiring require that firms demonstrate that hiring criteria which disadvantage minorities or women are related to subsequent job performance (U.S. Equal Employment Opportunity Commission, 1978).

The "individual characteristic" most often studied by economists (and often studied by others) is education. Attempts to use within-firm data to test whether more-educated individuals are more productive have produced very mixed results.

Berg (1970, Chapters 5-8) found more-educated workers no more productive in a variety of blue collar and (generally lower-level) white collar occupations. Swartz (1978, pp. 28-30) found education unrelated to secretarial performance at a large conglomerate firm. Brenner (1968, pp. 29-30) found high school record (grades, teachers' ratings, absenteeism) significantly related to job performance among Lockheed-California workers.

Supervisory and technical workers have been studied more extensively, but again without real consensus. Wagner (1960, p. 185) found years of schooling to be the best single predictor (among 31 variables) of the performance ratings of young executives. Mahoney *et al*. (1960, p. 156) found years of schooling "significantly" related (apparently at the 10 percent level) to supervisors' ratings in a sample of managers. However, Medoff

(1977, p. 23) and Medoff and Abraham (1978, p. 18) found education unrelated to performance but positively related to salary among managers at the same grade levels in three large corporations. Korman's (1968, p. 308) survey of studies using "personal history" data (including years of schooling) to predict managerial success concluded that "personal history data seem to have some predictive value for first-line supervisors" but were less useful for higher-level managers. Cambell _et al_. (1970, Chapter 8) reached more optimistic conclusions, but education was not among the personal history variables in several of the studies they reviewed.

Weisbrod and Karpoff (1968) found that, among AT&T college graduates, school quality and class rank were related to _salary_, which they used as a proxy for performance. Wise (1975) found both years of schooling and academic performance predictive of salary gains among Ford Motor Company engineers, and argued that these _gains_ reflected job performance. Kaufman (1978) found that graduate courses taken by engineers _after_ being hired were positively related to performance in research and development, but not in applied development or manufacturing.

An interesting body of evidence is evolving from employers' need to "validate" their hiring criteria (i.e., to prove that these criteria reflect differences in job performance) in order to satisfy federal standards for nondiscriminatory hiring; thus far, their efforts have not been very convincing (White and Francis, 1976).

While within-firm studies of worker performance can potentially provide valuable evidence on the question of whether individuals with more education (or more of some other characteristic) are more productive, they suffer from a potentially serious statistical flaw. Whether an employer should prefer more-educated applicants depends on whether the (potential) performance of such applicants is superior to that of less-educated applicants. Because potential performance of applicants not hired is not observed, the performance of those hired by the firm must be analyzed. Unfortunately, comparing performance of more- and less-educated workers is likely to give a biased reading of the difference in performance between more- and less-educated applicants. If firms prefer more-educated applicants, less-educated applicants who are hired must have some "compensating virtues" (Jencks *et al*., 1977, p. 183) known to the hirer but often not to the researcher. Thus, comparing performance of more- and less-educated workers understates the edge of the typical (i.e., randomly selected) more-educated worker over the typical less-educated worker. Moreover, the bias is more severe when the employer overestimates the importance of education (Brown, 1978).

This problem is not limited to research on schooling and job performance; it is present whenever one attempts to infer the determinants of performance when candidates have been selected using some information not available to the researcher and his computer (e.g., letters of reference). The issue has long been recognized in the psychometric literature on testing, but only highly restrictive models—in which selection is based solely on a weighted sum of known "test" scores—have been used successfully. The typical[2] real-world situation in which hiring is based on both recorded applicant characteristics and unrecorded, informal judgment is thought to present "an insuperable barrier to any analytic treatment" (Thorndike, 1949, p. 176).[3]

The purpose of this paper is to apply recent econometric research on "sample selection" problems to this employee-selection problem. In Section I, a model of selection and subsequent performance, and an estimation strategy, are presented. In Section II, the data--personnel files of applicants for supervisory positions at a nondurable manufacturing firm--are described. Empirical results are presented in Section III, and conclusions offered in Section IV.

I. Statistical Model

Let N be the number of applicants, indexed by i, n of whom are selected by the firm. Among applicants, performance Y_i depends on a set of applicant characteristics which are available to the researcher, X_i, and a random variable u_i which captures all other determinants of performance:

$$(1) \quad Y_i = X_i\beta + u_i \qquad\qquad i = 1, \ldots, N$$

The disturbance u_i is assumed to be normally distributed, with mean zero and constant variance σ^2 for all i; in particular, u_i is assumed to be uncorrelated with X_i among applicants.[4] Variables such as schooling, age, and previous experience would be elements of X_i; that portion of worker motivation that is uncorrelated with schooling, etc., is part of u_i. If Y_i and X_i are available for all applicants (or a random sample of applicants), one could estimate β by ordinary least squares.

As noted in the introduction, data on Y_i are typically unavailable for the $N-n$ applicants who are not hired, because Y is a measure of on-the-job performance. Unfortunately, these "missing" observations are not likely to be randomly selected, so that estimating equation (1) with the available n observations who were hired will be likely to produce biased estimates of β.

Presumably, firms hire those applicants who, based on available information, are thought to be most productive. The "available information" includes both data available to the researcher, Z,[5] and some data which the researcher cannot include in the estimation because it is not available (the selector's subjective impression of the applicant) or cannot be coded with confidence (letters of reference). Let e_i represent all the factors which affect applicant i's hiring decision that are uncorrelated with Z_i. The firm can be thought of as forming an estimate of the applicant's performance if hired, P_i, based on Z_i and e_i

$$(2) \quad P_i = Z_i\alpha + e_i \qquad\qquad i = 1, \ldots , N$$

and hires those with the highest P_i's (those with P_i greater than a cutoff score c). Assume that e_i is normally distributed with mean zero and constant variance for all i. Because the P_i's are not observed, we can "measure" them with arbitrary origin and scale; the normalizations $c = 0$, variance of $e = 1$ are the most convenient. Therefore, if $D_i = 1$ whenever individual i is hired, and zero otherwise, we have

$$(3) \quad \begin{aligned} D_i &= 1 \quad \text{iff} \quad Z_i\alpha + e_i > 0 \\ D_i &= 0 \quad \text{iff} \quad Z_i\alpha + e_i < 0 \end{aligned}$$

Equation (1) is often estimated by ordinary least squares from the subsample of individuals who were hired by the firm. To analyze the bias due to such sample selection, Heckman (1976, p. 477) rewrote equation (1) as

$$(1') \quad Y_i = X_i\beta + E(u_i|Z_i,D_i) + v_i$$

where v_i is the deviation of u_i from its conditional expectation. For the subsample of applicants who were hired, $D_i = 1$. From the assumption the u_i and e_i have a bivariate normal distribution, it follows that

$E(u_i|Z_i,D_i=1) = E(u_i|e_i > -Z_i\alpha) = \rho\sigma\lambda_i$, where ρ is the correlation of u and e (among applicants, $\lambda_i = f(Z_i\alpha)/F(Z_i\alpha)$, and $f(\cdot)$ and $F(\cdot)$ are the standard normal density- and distribution- functions.

Viewing λ_i as an omitted variable whose "true" coefficient is ρ , ordinary least squares is seen to produce biased estimates of β unless (i) λ_i and X_i are uncorrelated or (ii) $\rho = 0$. Since Z and X will typically have elements in common, condition (i) is unlikely to hold. The plausibility of condition (ii) depends on the selection process. If e_i reflects factors that are not related to performance (or if selection is conducted mechanically on the basis of Z_i , so that the variance of e is zero), e and u would be uncorrelated. More likely is the case where e reflects factors that are related to performance (e.g., the interviewer's estimate of the candidate's motivation), so that u and e are positively correlated. It seems likely that the more complete is the set of characteristics in X and Z , the smaller this correlation would be.

Griliches, Hall, and Hausman (1977, pp. 11-21) have presented a maximum-likelihood estimation procedure which incorporates Heckman's insight that the n observations on hired applicants must be "corrected" for the fact that they were not randomly selected. The data for the N applicants are arranged so that the n selected applicants are indexed by $i = 1, \ldots, n$ and the $(N-n)$ rejected applicants by $i = n + 1, \ldots, N$. The key to expressing the likelihood function is the string of equalities

$$(4)\quad \Pr(Z_i\alpha+e_i>0 \mid u_i) = \Pr(e_i>-Z_i\alpha \mid u_i) = \Pr\left[\frac{e_i-(\rho/\sigma)u_i}{(1-\rho^2)^{(1/2)}} > \frac{-Z_i\alpha-(\rho/\sigma)u_i}{(1-\rho^2)^{(1/2)}}\right]$$

$$= 1 - F\left[\frac{-Z_i\alpha-(\rho/\sigma)u_i}{(1-\rho^2)^{(1/2)}}\right] = F\left[\frac{Z_i\alpha+(\rho/\sigma)u_i}{(1-\rho^2)^{(1/2)}}\right]$$

Therefore, the likelihood function for an accepted applicant is

$$(5)\quad LF_i(D_i=1) = f(u_i/\sigma) \cdot Pr(Z_i\alpha+e_i>0 \mid u_i)$$
$$= (2\pi\sigma^2)^{-(1/2)} \exp\left[-(1/2)\left(\frac{Y_i-X_i\beta}{\sigma}\right)^2\right] F\left[\frac{Z_i\alpha+(\rho/\sigma)(Y_i-X_i\beta)}{(1-\rho^2)^{(1/2)}}\right]$$

The likelihood function for a rejected applicant is

$$(6)\quad LF_i(D_i=0) = Pr(Z_i\alpha+e_i<0) = F(-Z_i\alpha)$$

The logarithm of the likelihood function for the whole sample is

$$(7)\quad LLF = -(n/2)(\ell n\ 2\pi + \ell n\ \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-X_i\beta)^2$$
$$+ \sum_{i=1}^{n} \ell n\ F\left[\frac{Z_i\alpha+(\rho/\sigma)(Y_i-X_i\beta)}{(1-\rho^2)^{(1/2)}}\right] + \sum_{i=n+1}^{N} \ell n\ F(-Z_i\alpha)$$

Thus far, it has been implicitly assumed that performance ratings are available for all individuals chosen by the firm. In the employee-selection context, this assumption can be violated in two ways: some "selected" applicants will decide to reject the firm's offer of employment, and ratings may be unavailable for some of those hired. The best way to handle these exceptions depends on what is known about why they became exceptions (i.e., are they thought to be nonrandomly "selected" from among individuals selected by the firm?) and on their relative frequency in the data.[6] Because these exceptional cases are not very frequent in the data analyzed below, two simple expedients were considered. First, the exceptional cases were assumed to be randomly selected from among those who are chosen by the firm. If there are n individuals with Y_i available, m individuals selected by the firm but without ratings, and N-n-m individuals rejected by the firm, the last term of equation (7) is replaced by

$$\sum_{i=n+1}^{n+m} \ell n\ F(Z_i\alpha) + \sum_{i=n+m+1}^{N} \ell n\ F(-Z_i\alpha)\ .$$

Alternatively, we can reinterpret D_i to equal 1 whenever a rating is available, and zero otherwise—a more direct application of the "sample selection" model. Except as noted, the results in Section III are based on the first of these strategies.

II. Data

The data are taken from the personnel files of applicants for supervisory positions[7] at a medium-large (4000 employee) nondurable manufacturing plant. The firm's stated hiring standards were: "inside" applicants (candidates for promotion) must have a high school degree; "outside" applicants (new hires) must have a high school degree and previous supervisory experience, or have a college degree.

The files of 621 males who applied during 1968-70 and 1972-74 are available. Of these, 422 were either accepted or rejected by the firm (184 hired, 226 rejected, 12 offered jobs but rejected them). Of the remaining 198, 163 withdrew from consideration before a decision could be made, and no outcome was indicated for 35. Data for the 422 acted-upon applicants were used in this analysis.

The model described in Section I ignored applicants' reservation wages and thus is most plausible when the firm is constrained (or chooses) to pay each successful applicant the same wage. In fact, this was approximately true for this sample. Starting salary data revealed a clear tendency for new supervisors to receive the "going" monthly rate at time of hire. By plotting starting salaries chronologically, a starting rate which changed about once per year over the period studied could be identified; 78 per cent of those hired received this starting rate.

The rating used to estimate equation (1) was the individual's latest rating as a first-line supervisor. Ratings were available for 161 of the 184 hired supervisors. The firm used two different rating forms. Form 1, used prior to 1974, had four ratings (outstanding, average, satisfactory, below average); Form 2, used since 1974, had five ratings (excels, meets requirements, above minimum, meets minimum, below minimum). To make the ratings comparable, each was "Z-scored," using a procedure described by Medoff (1977) and Freeman (1978, p. 135). It was assumed that an underlying continuous performance variable w has a standard-normal distribution. The frequency distribution of the performance categories for all ratings was tabulated. From these distributions, the value of the standard-normal variate at the category boundaries, and the expected value of the standard-normal variate within each category can be computed.[8] For example, suppose 10 percent of all ratings fall in the lowest category. From a standard-normal cumulative distribution table, we can infer that the lowest rating category corresponds to $w < -1.28$, since $\Pr(w < -1.28) = .10$. The expected value of w within the lowest rating category is $E(w|w < -1.28) = -1.75$. If 20 percent of ratings fall in the next lowest category, this category corresponds to $-1.28 < w < -.52$, since $\Pr(w < -.52) = .30$, and $E(w|-1.28 < w < -.52) = -.86$.

The information available from the personnel files is summarized in Table 1. These variables represent most of the factors generally thought to be important in hiring decisions at this level; in any case, they represent nearly all of the information that was routinely recorded at this plant during the hiring process.[9]

In the estimation of the model, Year of Application is assumed to influence selection but not performance, while Years in Supervisory Position and Rating is Form 2 affect the performance rating but not selection.

Table 1.

Means of Variables

Variable	Mean (S.D.)
1. Schooling (years completed)	13.6 (2.11)
2. Potential Experience (age-schooling-5)	9.91 (7.81)
3. Previous Supervisory Experience (1 = yes; 0 = no)	.453
4. Inside Applicants (1 = yes; 0 = no)	.400
5. Tenure (months with firm at time of hiring decision/12)	6.22 (5.91)
6. Married (1 = married, spouse present; 0 = all other)	.713
7. Honorable Discharge (1 = honorable discharge; 0 = all other)	.379
8. In Reserves (1 = in reserves; 0 = all other)	.076
9. Year of Hiring Decision (year + month/12 - 1970)	2.65 (1.86)
10. Selected (1 = hired or offer rejected; 0 = rejected)	.464
11. Rating is Form 2 (1 = yes; 0 = no)	.913
12. Years in Supervisory Position (rating date - decision date, in years)	2.95 (1.89)
13. Rating (Z-scored)	.065 (.833)

N = 422

Notes:
1. Means and standard deviations on lines 5 and 11-13 are means among inside applicants and rated workers, respectively.
2. Missing values were coded at sample means for variables 1, 2, and 5, and included in "no" category for variables 6-8. Number of missing values, by variable: Schooling(5), Potential Experience(16), Tenure(1), Married(11), Honorable Discharge and In Reserves (38).

The remaining determinants of selection are assumed affect performance, and vice versa. On a priori grounds, one expects those characteristics which are preferred in the selection process to have positive effects on performance. While not strictly necessary for identification, the availability of some variables not entering both equations should alleviate reservations about estimating sample selection models in other contexts (Welch, 1977, p. 455).

The most important aspect of the data, however, is the availability of data on rejected applicants' Z's. Thus, we have a "censored" sample rather than the "truncated" samples (Heckman, 1976, p. 476) used in the employee-performance studies cited in the introduction.

III. Estimates

Ordinary least squares estimates of the performance equation are presented in the first column of Table 2. Schooling and Potential Experience have positive coefficients, though neither is very large (a year of schooling corresponding to a .05 standard deviation improvement in the rating) nor statistically significant. Previous Supervisory Experience has a negative coefficient with a large standard error. Neither Inside Applicant nor Tenure is significant, and the estimated advantage of an inside applicant with average tenure over an outside applicant (-.017 + .009*6.22 = .04) is small. Married individuals, those with honorable discharges, and those in the reserves receive lower ratings. Years in Supervisory Position is strongly and positively related to the performance rating.

The results are similar to what one might expect from ordinary least squares estimates in the face of sample selection: some variables which might be expected to have positive effects (Schooling, Potential Experience) have negligible coefficients, while others whose impact on performance might be doubted are strongly negative (Married, Honorable Discharge).[10] Moreover, the variable which the firm cannot use for selection—Years as a Supervisor (with <u>this</u> firm)--is positive and statistically significant.

The maximum-likelihood estimates of both selection and performance equations are also presented in Table 2. The selection equation (rightmost column) is of considerable interest. It indicates that, contrary to the stated hiring policy, Schooling is only weakly related to the probability

Table 2.

Estimates of Equations (1) and (2)

Variables	Performance: Equation (1) OLS	Performance: Equation (1) ML	Selection: Equation (2) ML
Constant	-.871 (.833)	-1.420 (.926)	-1.209 (.885)
Schooling	.043 (.048)	.050 (.050)	.045 (.054)
Potential Experience	.003 (.013)	.005 (.014)	.009 (.013)
Previous Super. Exper.	-.155 (.175)	-.246 (.193)	-.337* (.164)
Inside Applicant	.017 (.207)	-.037 (.205)	-.493 (.254)
Tenure	.009 (.017)	.025 (.020)	.147* (.048)
Married	-.093 (.203)	.081 (.292)	.759* (.174)
Honorable Discharge	-.243 (.150)	-.126 (.202)	.689* (.159)
In Reserves	-.144 (.219)	.027 (.305)	.932* (.260)
Year of Hiring Decision			-.161* (.041)
Rating is Form 2	.315 (.243)	.197 (.293)	
Years in Super. Position	.088* (.036)	.115* (.047)	
R^2	.100	.101	

Standard errors in parentheses beneath coefficients

* = "significant" (t-ratio greater than 1.96)

of selection, and Previous Supervisory Experience has a significant negative effect. Potential Experience has a very small effect, while inside applicants,[11] married workers, and applicants with honorable discharges or reservist status are preferred. Finally, Year of Decision is significant and negative, indicating that standards were rising over time.

A comparison of the maximum-likelihood estimates of the performance equation (middle column) with the ordinary least squares estimates reveals predictable changes in the coefficients. The estimated coefficients of Schooling and Potential Experience change very little, while Previous Supervisory Experience becomes more negative. The coefficients of the other variables which were important in selection become more positive (the average inside applicant's advantage rises from .04 to .12), but while the changes are fairly sizeable, none of these coefficients becomes "significant". In general, the standard errors of these coefficients also increase considerably. Years in Supervisory Position remains significantly related to performance.

In addition to the changes in coefficients, the impact of accounting for selection on the ability to predict performance is of interest. The "R^2" presented for the maximum likelihood estimate of the performance equation uses equation (1') to create a measure analogous to R^2 from ordinary least squares. Using the maximum likelihood estimates of α, β, ρ, and σ, a predicted value $X_i\hat{\beta} + \hat{\rho}\hat{\sigma}\hat{\lambda}_i$ was calculated for each rated individual. The squared correlation of this constructed variable with the actual performance rating was .101, a very small improvement over the ordinary least squares value.

As shown in Section I, the bias in ordinary least squares estimates (and hence the anticipated changes from maximum likelihood estimates) depends on the correlation between the error terms in the selection and performance equations. For the equations in Table 2, the maximum-likelihood estimate of this correlation was 0.43 (the hypothesis $\rho = 0$ could not be rejected).

The sensitivity of the results to details of the specification was examined by several exeriments, two of which are reported in Table 3. Because the selection equations were virtually unaffected, only the alternative estimates of the performance equation will be discussed.

The use of the last rating as a measure of performance could be challenged on the grounds that it obscures the value to the firm of more-qualified applicants. Those who perform poorly initially might be given extra attention by their supervisors and eventually perform as well as those who were better qualified initially. To test this conjecture, the first two columns of Table 3 report OLS and ML estimates of the performance equation when the _first_ available rating for each individual is used to measure performance.[12] The OLS coefficients display the same patterns as those in Table 3, except that Potential Experience has a much larger coefficient which verges on statistical significance. In general, the ML estimates differ from their OLS counterparts less than when the last rating was used; this reflects the weaker correlation between the selection and performance equations' disturbances here compared with Table 2. (.21 vs. .43). The effect of years in Supervisory Position with this firm is nearly three times as large as in Table 2; this suggests that the marginal effect of such ex-

Table 3.

Estimates of Equations (1) and (2)

	First Rating			Last Rating (half-normal)		
	Performance		Selection	Performance		Selection
Variable	OLS	ML	ML	OLS	ML	ML
Constant	-.087 (.797)	-.331 (1.063)	1.151 (.904)	.204 (.500	-.499 (.529)	-1.293 (.846)
Schooling	-.025 (.049)	-.023 (.044)	.042 (.055)	.027 (.029)	.036 (.032)	.047 (.052)
Potential Experience	.024 (.013)	.025* (.012)	.007 (.013)	.003 (.008)	.005 (.008)	.009 (.013)
Previous Super. Exper.	-.001 (.173)	-.043 (.236)	-.342* (.162)	-.064 (.105)	-.199 (.108)	-.311 (.165)
Inside Applicant	.178 (.210)	.152 (.247)	-.480 (.246)	.040 (.125)	.043 (.133)	-.442 (.249)
Tenure	-.000 (.017)	.006 (.033)	.146* (.048)	.005 (.010)	.026* (.012)	.136* (.046)
Married	-.225 (.204)	-.145 (.351)	.761* (.176)	-.038 (.122)	.191 (.160)	.765* (.174)
Honorable Discharge	-.291 (.151)	-.237 (.254)	.695* (.160)	-.132 (.090)	.032 (.134)	.673* (.157)
In Reserves	.030 (.220)	.110 (.364)	.958* (.261)	-.094 (.132)	.145 (.186)	.886* (.264)
Year of Hiring Decision			-.168* (.041)			-.148* (.040)
Rating is Form 2	-.263 (.140)	-.307 (.208)		.194 (.146)	.011 (.176	
Years in Super. Position	.237* (.074)	.249* (.071)		.050* (.022)	.089* (.026)	
R^2	.168	.171		.087	.082	

Standard errors in parentheses beneath coefficients.

* = "significant" (t-ratio greater than 1.96)

perience dies out rather quickly. (The mean for this variable at first rating was 0.96 years compared with 2.95 years at last rating.)

In summarizing the ratings in a single score, the z-scoring procedure assumed that the underlying continuous performance variable had a normal distribution. A plausible objection is that, since selection truncates this distribution, the distribution of observed performance should be asymmetric, with (much of) the lower tail removed. To test the sensitivity of the results to the distribution of performance among those rated assumed in z-scoring, an extreme alternative was considered. Since roughly half of the applicants were hired, it was assumed that performance followed the upper half of a normal distribution, i.e., $f(y) = (2/\pi)^{\frac{1}{2}} \exp(-y^2/2) \quad 0 < y < +\infty$, and the rating categories were z-scored on this basis. As the last two columns of Table 3 demonstrate, the most important effect of this respecification is the improvement in the coefficients of marital and military status variables. None, however, is statistically significant. The improvement is due to the increased estimated correlation between selection and performance disturbances, which now reaches .83.

Two further experiments were considered. First, a dummy variable for those hired at premium pay was added to the performance equation using the Table 2 specifications. This variable was statistically insignificant, and the other coefficients were not appreciably affected. Second, the variable D_i in equation (3) was redefined as "rated" rather than "selected", as discussed in Section I. Once again, this modification had no appreciable affect on the performance equation.

IV Conclusions

Ordinary least squares estimates of the determinants of employee performance are likely to give biased estimates of the importance of the various factors which firms use in hiring. The magnitude of this bias, however, depends on the selection process of the firm, being most important when the firm accurately judges unrecorded attributes of applicants which are in fact important determinants of later performance. Consequently, the magnitude of such biases will vary from one context to another—as will the determinants of performance themselves.

For the firm studied here maximum likelihood estimates of the determinants of employee performance differed appreciably from ordinary least squares estimates in some cases, suggesting that selection bias of some importance. However, none of the applicant characteristics available at time of hiring were significantly related to later performance. In some cases (Schooling, Potential Experience) these variables appear to have had little impact on the hiring decision, so that their lack of impact on performance is unremarkable. The remaining variables were, however, important in the decision procedure.

Years in Supervisory Position was strongly related to performance. While this appears to contradict the findings of Medoff (1977) and Medoff and Abraham (1978), such a conclusion would be unwarranted. In each of the samples they studied, the average level of company experience (presumably, mostly supervisory) was 12 years or more, compared with three years of supervisory experience with the firm in the sample studied here. Moreover, the coefficient was much smaller for last rating than first rating, suggesting that the marginal effect of such experience after a year or two could be very small.

The apparent lack of relationship between the characteristics preferred by the firm and subsequent performance is subject to two qualifications. First, characteristics such as being married may be used by the firm to select those who will stay with the firm (minimizing future hiring and training costs) even if they perform no better on the job. Second, the parameters of the performance equation are not estimated with great precision, due in part to smaller sample size than previous applications of the sample selection model. If this finding is supported by later research, it would mean that the model could be used successfully only for larger firms or less elite jobs than studied here. However, the difference in estimated coefficients due to correcting for selection bias reported above suggests that this problem may be of considerable practical importance in some contexts.

Footnotes

1. Cost-minimizing behavior is a major point of agreement for "human-capital" (Becker, 1975; Mincer, 1974) and "screening" (Arrow, 1973; Spence, 1973) models of schooling.

2. Campbell *et al*. (1969, p. 39) state that all of the sample of large firms they studied selected managers in this way.

3. The sample-selection problem is treated as a "restriction in range" in this literature, which emphasizes correlations rather than regression coefficients. More recent studies give no indication that the "insuperable barrier" has been overcome; e.g., Whitla (1968, p. 470) who reports that the restrictive models "seem to have fallen into general disfavor" because their assumptions "are often hard to meet", Schmidt *et al*. (1976) who discuss the importance of selection in small-sample studies, and Maxwell (1974, p. 59) who presents a restriction in range correction for regression coefficients.

4. This implies that we "give credit" to X for any performance determinant which is correlated with X. This is appropriate when the firm's goal is to *predict* performance, rather than to determine its "causes."

5. Presumably, elements of Z are also elements of X. However, X may include variables not available at time of hiring, but subsequently available for those who are hired (e.g., time between selection and measurement of performance).

6. The relative frequency affects the ability of the data to "support" a more complicated analysis as well as the likely practical value of undertaking it.

7. Twenty-one "supervisory" applicants were hired into nonsupervisory (generally technical) positions and were deleted from the analysis.

8. Guilford (1954, pp. 181-2) discusses a similar procedure, except that the conditional median is used instead of the conditional mean.

9. Three available variables were excluded to limit the number of parameters to be estimated; they were nearly uncorrelated with other variables among applicants (dummy variables for health problems, G.E.D. high school degree, and race).

10. If the firm prefers, say, married applicants, one would anticipate that the estimated coefficient of Married would be negatively biased. If the true coefficient were zero, the estimated coefficient should be negative.

11. The advantage of the average inside applicant is $-.493 + .147 * 6.22 = .42$.

12. Using the average rating produced results quite similar to those for first rating, presented below.

References

Arrow, Kenneth, "Higher Education as a Filter," Journal of Public Economics, vol. 2, no. 3, July 1973, pp. 193-216.

Becker, Gary, Human Capital (New York: National Bureau of Economic Research, second edition, 1975).

Berg, Ivar, Education and Jobs (Boston: Beacon Press, 1970).

Brenner, Marshall H., "Use of High School Data to Predict Work Performance," Journal of Applied Psychology, vol. 52, no. 1, February 1968, pp. 29-30.

Brown, Charles, "Education and Jobs: An Interpretation," Journal of Human Resources, vol. 13, no. 3, Summer 1978, pp. 416-421.

Campbell, John P. et al. Managerial Behavior, Performance, and Effectiveness (New York: McGraw-Hill, 1970).

Griliches, Zvi; Hall, Bronwyn; and Hausman, Jerry, "Missing Data and Self-Selection in Large Panels," presented at the INSEE Conference on Economics of Panel Data, August 1977.

Guilford, J. P. Psychometric Methods (New York: McGraw-Hill, 1954).

Heckman, James, "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables, and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, vol. 5, no. 4, November 1970, pp. 475-492.

Kaufman, H. H., "Continuing Education and Job Performance: A Longitudinal Study," Journal of Applied Psychology, vol. 63, no. 2, April 1978, pp. 248-251.

Korman, Abraham K., "The Prediction of Managerial Performance: A Review," Personnel Psychology, vol. 21, no. 3, Autumn 1968, pp. 295-322.

Mahoney, Thomas A., et al., "Predicting Managerial Effectiveness," Personnel Psychology, vol. 13, no. 2, Summer 1960, pp. 147-163.

Maxwell, A. E., "Limitations on the Use of the Multiple Linear Regression Model," British Journal of Mathematical and Statistical Psychology, vol. 27, part 2, November 1974, pp. 51-62.

Mincer, Jacob. Schooling, Experience, and Earnings (New York: National Bureau of Economic Research, 1974).

Medoff, James, "The Earnings Function: A Glimpse Inside the Black Box," Discussion Paper No. 594, Harvard Institute of Economic Research, 1977.

Medoff, James and Abraham, Katherine, "Experience, Performance, and Earnings," xeroxed, May 1978.

Schmidt, Frank L.; Hunter, John E.; and Urry, Vern W., "Statistical Power in Criterion-Related Validation Studies," Journal of Applied Psychology, vol. 61, no. 4, 1976, pp. 473-485.

Spence, Michael, "Job Market Signalling," Quarterly Journal of Economics, vol. 87, no. 3, August 1973, pp. 355-374.

Swartz, Katherine, "Screening and the Hiring Process," Working Paper No. 77-9, University of Maryland Department of Economics and Bureau of Business and Economic Research, 1978.

Jencks, Christopher, et al. Inequality (New York: Basic Books, 1972).

Thorndike, Robert. Personnel Selection (New York: John Wiley and Sons, 1949).

U.S. Equal Employment Opportunity Commission, "Uniform Guidelines on Employee Selection Procedures," 1978.

Wagner, Edwin E., "Predicting Success for Young Executives from Objective Test Scores and Personal Data," Personnel Psychology, vol. 13, no. 2, Summer 1960, pp. 181-186.

Welch, Finis, "What Have We Learned from Empirical Studies of Unemployment Insurance?", Industrial and Labor Relations Review, vol. 30, no. 4, July 1977, pp. 451-462.

White, David, and Francis, Richard, "Title VII and the Masters of Reality: Eliminating Credentialism in the American Labor Market," Georgetown Law Journal, vol. 64, no. 6, July 1976, pp. 1213-1244.

Whitla, Dean K., ed., Handbook of Measurement and Assessment in Behavioral Sciences (Reading, MA: Addison-Wesley, 1968).

Wise, David, "Academic Achievement and Job Performance," American Economic Review, vol. 65, no. 3, June 1975, pp. 350-366.