Reexamining the Returns to Training

Functional Form, Magnitude, and Interpretation

Harley Frazis
Mark A. Loewenstein

ABSTRACT

We investigate the functional form for formal training in a wage equation and derive estimates of its rate of return. The cube root fits best in our two data sets. We show that if wages are not adjusted continuously, estimating the return to training requires one lag and one lead of training. Using the cube root and a semi-nonparametric estimator, estimated returns are 150–180 percent. Adjusting for heterogeneity in wage growth, promotions, and direct costs reduces the return to 40–50 percent. We find evidence of heterogeneity in returns. Our estimates can thus be regarded as the return to training for the trained, but cannot be extrapolated to the untrained.

I. Introduction

In recent years, a substantial literature analyzing the extent and consequences of on-the-job training has emerged, taking advantage of new data sets with direct measures of training.[1] Studies find support for the human capital model's prediction that a worker's wage is positively related to past investments in his training. Barron, Black, and Loewenstein (1989) note that "training is one of only a few variables affecting wage and productivity growth."

Both authors are researchers at the Bureau of Labor Statistics. *The authors thank Dan Black, Marilyn Manser, Terra McKinnish, participants in a session of the 1998 Southern Economic Association meetings and at seminars at Virginia Commonwealth University and George Washington University, and an anonymous referee for helpful comments. We thank Alexander Eidelman for assistance in constructing the NLSY data set, Hector Rodriguez for help with figures, and James Spletzer both for help with the NLSY data and for comments. The views expressed here are those of the authors and do not necessarily reflect the views of the U.S. Department of Labor or the Bureau of Labor Statistics. The data used in this article can be obtained beginning October 2005 through September 2008 from Harley Frazis, Bureau of Labor Statistics, 2 Massachusetts Avenue, NE, Room 4945, Washington, D.C. 20212-0001, e-mail Frazis.Harley@bls.gov.*
[Submitted July 2001; accepted February 2004]
ISSN 022-166X E-ISSN 1548-8004

1. An extensive but not exhaustive list of references is found in our longer working paper, Frazis and Loewenstein (2003).

However, in many respects the literature on training lags behind the more developed literature on the returns to schooling. While studies of the rate of return to schooling are numerous, we are aware of few studies that attempt to estimate rates of return to training.[2] At a more basic empirical level, while the best simple functional form to characterize the earnings-schooling relationship has been settled since Heckman and Polachek (1974), researchers have paid little attention to the choice of the appropriate functional form for the earnings-training relationship. Differences in functional forms across studies make comparisons difficult. This difficulty is compounded by the fact that researchers using different functional forms have tended to use different data sets: while users of the Employer Opportunity Pilot Project (EOPP) data and the closely related Small Business Administration (SBA) data have generally used log specifications,[3] researchers using the National Longitudinal Study of Youth (NLSY) have used linear specifications or specifications that estimate the return to a spell of training without making use of information on the duration of the spell.[4]

This paper has three goals. First, we seek to perform a service similar to Heckman and Polachek (1974) by investigating the choice of the appropriate functional form for formal training in a wage equation. Second, we derive estimates of the rate of return to formal training. We use both our preferred functional form and, where possible, a semi-nonparametric estimator to derive rate-of-return estimates. Third, because estimated returns from our fixed-effect regressions are quite high, we consider possible explanations for these apparently high returns.

II. Functional Form of the Training-Wages Relationship

In this section, we compare several different simple functional forms to determine which best describes the relationship between formal training and log wages. We use the NLSY and EOPP data sets to confirm the robustness of our results.

A. NLSY Data

The NLSY is a data set of 12,686 individuals who were aged 14 to 21 in 1979. These youth were interviewed annually from 1979 to 1994, and every two years since then. Response rates were more than 90 percent for each year until 1996, and as of 2000 were 83 percent. We use data from the 1979 through 2000 surveys.[5]

The training section of the survey begins with the question, "Since [the date of the

2. Mincer's (1989) review article in *Education Researcher* calculates rates of return in the range of 32–48 percent before depreciation. Bartel (1995), using a company data set, estimates the rate of return to training at 58 percent before depreciation; her calculation includes direct costs of training. Allowing for depreciation substantially reduces these numbers—Mincer's range after correction is from 4 to 26 percent, using Lillard and Tan's (1986) estimated 15–20 percent depreciation rate; Bartel's is 42 percent with 10 percent depreciation and 26 percent with 20 percent deprecation. Interestingly, Lengermann (1999) finds no evidence that the return to training depreciates with time.

3. For example, see Barron, Black, and Loewenstein (1989) and Barron, Berger, and Black (1997).

4. Lynch (1992) and Parent (1999) use linear specifications while Loewenstein and Spletzer (1996) and Lengermann (1999) use spells.

5. Individuals were not interviewed in 1995. From 1996 on, the survey is being conducted every other year.

last interview], did you attend any training program or any on-the-job training designed to help people find a job, improve job skills, or learn a new job?" Individuals who answer yes to this question are then asked a series of detailed questions about each of their different training spells. In 1988 and thereafter, individuals are asked about the duration of their various training spells in weeks and the average number of hours each week that were spent in training. For each training spell in a given year, we have calculated the number of hours spent in training as the product of the duration in weeks and the average number of hours spent in training during a week.

The training questions were changed somewhat in 1988. From 1979–86, detailed information was obtained only on training spells that lasted longer than one month.[6] We have used the information contained in the later surveys to impute hours spent in training for training spells in the early surveys that last less than one month.[7]

The focus of our analysis is training whose explicit cost is at least partly paid for by the employer.[8] Information on who paid for training is available only after 1987; prior to 1987, we include only company training and spells less than one month. (The post 1987 data indicate that company training was generally paid for by the employer. Prior to 1987, individuals with spells less than one month were not asked about the type of training they received; the post 1987 data indicate that short spells are generally employer-paid.) Training not paid for by the employer (after 1987) and noncompany training (prior to 1987) are accumulated in a separate variable.

In investigating the effect of training on wages, it is important to distinguish between training that took place on the current job and training that took place on other jobs. By comparing the beginning and ending dates of a training spell with the date that the individual started working at his current job, we are able to classify a training spell as occurring on the current job or on a previous job.[9] When there is some ambiguity as to whether training occurred on the current job or in a previous job, we classify the training as occurring in the current job. Our results are not sensitive to this choice. The key training variable used in the empirical work to follow is total accumulated completed training on the current job.

B. Basic Results

Our basic specification is:

$$(1) \quad \ln W_{ijt} = X_{ijt}\beta_1 + f(T)\beta_2 + \alpha_i + \theta_{ij} + \omega_t + \varepsilon_{ijt}$$

for person *i* in job *j* at time *t*, where *W* is the wage rate, *X* is a vector of time-varying control variables, *T* is hours of training on the current job, $f(\cdot)$ varies by specification, α_i and θ_{ij} are permanent person and job-match specific error terms, ω_t is a year effect, and ε_{ijt} is a mean zero error term, uncorrelated with X_{ijt}. All specifications are run as

6. Training questions were not asked in 1987.

7. More details about the imputation procedure are available in Frazis and Loewenstein (2003).

8. Naturally, this cost can be passed on to the worker in the form of a lower wage. Eighty-five percent of the training spells in the NLS are at least partially paid for by the employer. We focus on this training because it would appear to correspond most closely to the on-the-job training concept referred to by Becker and subsequent human capital theorists.

9. In cases where the individual holds more than one job simultaneously, we assume that training occurs on the individual's main job.

fixed-effect regressions within jobs. As the fixed effect will absorb both unchanging individual characteristics and job characteristics, the X vector is mostly comprised of functions of tenure and interactions of tenure with other variables. Specifically, the X vector consists of tenure, tenure squared, tenure cubed, and interactions of the three tenure terms with age at start of job, experience at start of job, AFQT,[10] years of education, ever married, part-time, union, two dummies for initial occupation in the job, Black, Hispanic, female, enrolled in school, and missing value indicators for AFQT, union, and part-time. Years of education (which occasionally changes within a job), nonemployer-paid training,[11] and dummies for ever married, part-time, enrolled in school, missing part-time, and year dummies are also included. As an additional control for training, we include a count of spells with missing training duration (most of these occur before 1988).

We exclude observations with missing values on variables other than AFQT, union, and part-time. We also exclude observations with real wages below \$1 or above \$100 in 1982–84 dollars, or with log wages where the absolute value of the difference with the job mean is greater than 1.5 (which is a little more than 7.5 standard deviations). Finally, we exclude the military subsample, and jobs where for half or more observations on that job the respondent is an active member of the armed forces, self-employed, in a farm occupation, or enrolled in school at any time between interviews. The resulting sample has 75,698 observations from 17,809 jobs.

Descriptive statistics are shown in Table 1. We note that the standard deviation of training conditional on receiving any training is larger than the mean; the positive training distribution is quite skewed to the right. Log-normality appears to be a good approximation.

As an aid to determining the best simple functional form, we use the Box-Cox transformation (Greene 2000) by first estimating a model where $f(T) = B(T; \lambda) \equiv (T^{\lambda} - 1)/\lambda$. Values of λ of one and zero correspond to $f(T) = T$ and $f(T) = \ln(T)$, respectively. The estimation is done by nonlinear least squares. The estimated value of λ of 0.350 (bootstrap standard error 0.062) is very close to the value of one-third corresponding to the cube root.

The results for different functional forms for training are shown in Table 2.[12] In addition to the cube root, we include linear, log, and quadratic specifications, a specification where the training variable is a dummy indicating whether any training has been completed on the current job, and a specification with both this dummy and a linear term. In the "log" specification, $f(T) = \ln(T + 1)$, where T is number of hours of training. The table shows $\overline{R}^2$s (explained variance as a proportion of within-job variance) and the total effect at the median number of hours of training (60), where the median is calculated across all observations with a positive stock of training on the current job. The differences in fit appear slight. However, the best-fitting specification—the cube root—increases $\overline{R}^2$ more than twice as much as the worst-fitting specification—the linear—relative to the fit excluding training variables. The quadratic specification and the dummy specification are little improvement on the linear, while

10. Specifically, the residual from a regression of AFQT on dummies for year of birth.

11. A preliminary functional form analysis showed that a linear specification was best for nonemployer-paid training.

12. Estimating regressions for men and women separately gives results quite similar to Table 2.

Table 1
Descriptive Statistics, NLSY

Variable	Mean	Standard Deviation	Minimum	Maximum
Ln wage	1.88	0.49	0.00	4.53
Number of training spells, current job	0.54	1.34	0.00	21.00
Training hours	40.60	209.14	0.00	9,260.00
Ln (training + 1)	1.02	1.90	0.00	9.13
Training hours, training > 0	164.60	396.11	0.50	9,260.00
Ln (training + 1), training > 0	4.12	1.39	0.41	9.13
Years tenure	3.80	3.68	0.00	22.77
Black	0.26	0.44	0.00	1.00
Hispanic	0.18	0.38	0.00	1.00
Age at start of job	25.28	4.78	14.90	41.19
Years experience at start of job	5.88	4.07	0.00	23.47
Female	0.50	0.50	0.00	1.00
AFQT (residual)	0.37	20.17	–65.48	45.94
Years education	12.77	2.27	0.00	20.00
Ever married	0.63	0.48	0.00	1.00
Union	0.20	0.40	0.00	1.00
Managerial/professional (first year in job)	0.17	0.38	0.00	1.00
Other white-collar (first year in job)	0.34	0.47	0.00	1.00
Part-time	0.12	0.32	0.00	1.00
Hours outside training on current job	18.47	165.59	0.00	5,440.00
Missing AFQT	0.06	0.23	0.00	1.00
Missing union	0.05	0.21	0.00	1.00
Missing part-time	0.00	0.03	0.00	1.00
Number of spells missing training hours, current job	0.01	0.12	0.00	4.00
Ever training spell, current employer	0.26	0.44	0.00	1.00
Any training spell, current period	0.10	0.30	0.00	1.00
Number of jobs	17,809			
Observations	75,698			

Table 2
Returns to Training for Different Functional Forms, NLSY

Specification	$\bar{R}^2$	Fraction Fourier Series Explained	Total Effect at Median
Complete sample			
No training variables	0.2033	—	—
Dummy	0.2042	0.624	0.031
Linear	0.2040	0.332	0.003
Quadratic	0.2042	0.461	0.005
Cube root	0.2050	0.842	0.036
Log	0.2049	0.822	0.041
Dummy + linear	0.2047	0.732	0.029
Fourier series	0.2057	—	0.039
Number of jobs	17,809		
Observations	75,698		
Training Outliers Omitted*			
No training variables	0.2023	—	—
Dummy	0.2032	0.630	0.031
Linear	0.2031	0.481	0.007
Quadratic	0.2035	0.716	0.014
Cube root	0.2040	0.842	0.037
Log	0.2038	0.823	0.041
Dummy + linear	0.2037	0.762	0.029
Fourier series	0.2047	—	0.034
Number of jobs	17,788		
Observations	75,497		

*Top 1 percent of training duration omitted.

the log specification is close to the cube root. The dummy-plus-linear specification has a somewhat lower fit than the log and the cube root. The results indicate that the effect of training on wages is highly nonlinear, with the effect declining more rapidly than implied by a quadratic specification. There is no evidence that the presence of an incidence effect explains the nonlinearity.[13]

To contrast the effects of training on wages implied by the different functional forms, the last column in Table 2 shows the predicted effect of training at the median of the distribution of positive hours of training. The implied effect of the median hours of training differs by more than a factor of 12 between the different specifications. The log specification shows the largest effect, more than 4 percent, with the cube root yielding a slightly smaller effect. The linear and quadratic specifications apparently greatly understate the impact of training on wages.

13. Adding a dummy term to the cube root specification yields a dummy coefficient that is negative, small in magnitude, and not significant.

One might suspect that the better fit of the log and cube-root specifications simply reflects the fact that these functions' compression of the right tail of the training distribution reduces the influence of outliers. To test for this, we omitted the top one percent of the distribution of positive training. The total effect of the median amount of positive training increases for the linear and quadratic specifications, but is still far below the other specifications. The cube root specification is still the best fitting; excepting perhaps the quadratic, there is no marked improvement in the fit of the other specifications relative to the cube root.

All of the above specifications are parsimonious, with the rate of decline determined by the functional form. To compare the patterns of returns implied by these specifications with those obtained from less restrictive specifications, we use a seminonparametric estimator: the Fourier series expansion (Gallant 1981). A *K*th order Fourier series is a linear combination of cosine and sine terms, or $f^*(T) = \sum_{j=1}^{K}(\alpha_{1j} \cos(jT) + \alpha_{2j}\sin(jT))$. A function's Fourier expansion has the property that the differences between the value of a function f and the value of its Fourier expansion f^* and between the derivatives of f and the derivatives of f^* can be minimized to an arbitrary degree over the range of the function by choosing K to be sufficiently large. It thus provides a global approximation to the true function, rather than a local approximation (as in a Taylor series expansion).

In practice, linear and quadratic terms are usually added to the expansion. Moreover, for nonperiodic functions the variable T needs to be transformed to a variable T^* such that $0 < T^* < 2\pi$, after which the expansion can be implemented as:

$$f^*(T^*) = \delta_1 T^* + \delta_2 T^{*2} + \sum_{j=1}^{K}(\alpha_{1j}\cos(jT^*) + \alpha_{2j}\sin(jT^*)).$$

In our case, due to the essentially log-normal distribution of training, it is computationally convenient to work with the log of training as the basis for the Fourier expansion. We thus adopt the transformation $T^* = 0.001 + c\ln(T+1)$, with c chosen so that the maximum value of T^* is close to 2π. We chose K to minimize the sum of squared prediction errors $CV = \sum_i(y_i - x_{iK}\hat{\beta}_{-iK})^2$, where X_{iK} is the complete vector of regressors for the Kth order expansion and $\hat{\beta}_{-iK}$ is the corresponding coefficient vector from a regression omitting observation i. Andrews (1991) shows this criterion is asymptotically optimal in the sense that the probability of choosing the K that minimizes the expected sum of squared errors converges to one as the sample size increases, even in the presence of heteroscedasticity.[14] We searched all orders of the expansion from $K = 1$ to 14. The order K is 13 for both the complete and outlier-omitted sample.

We calculate the statistic $Q^2 \equiv 1 - \frac{\sum(f(T) - f^*(T^*))^2}{\sum(f^*(T^*) - 0)^2}$ to obtain a convenient summary measure of the closeness of fit between an arbitrary specification $f(T)$ and the estimated Fourier series $f^*(T^*)$.[15] Analogous to the traditional R^2, which measures the percentage reduction in the sum of the squared distance between the dependent variable and the predicted value relative to a model with only a constant, Q^2 measures the percentage reduction in the squared distance between the Fourier

14. However, asymptotic optimality is proven only when observations are independent.
15. We are grateful to Dan Black for suggesting this type of statistic.

series and $f(T)$ relative to a specification which omits training. As can be seen in the third column of Table 2, the cube root specification is closest to the Fourier series, and the linear specification is the furthest. Indeed, the cube root specification explains more than 80 percent of the squared distance between the Fourier series and a specification without training, while the linear specification explains only 33–48 percent depending on the sample.

Figure 1 plots the effect of training estimated in the sample without outliers for all specifications except for the dummy specification. The effect is plotted against log training since a linear scale would overly compress the range where the data are concentrated. The range of the figure is further restricted to the 5th through 95th percentile of the positive training distribution. The volatility of the Fourier series apparent in the figure suggests that much of the variation in the Fourier function unexplained by the better-fitting functional forms is spurious. Consistent with the Q^2 statistics, the figure shows that the linear and quadratic specifications fit the basic pattern of returns in the Fourier series expansion worse than the other functional forms over most of the range of the data, especially between the 25th and 75th percentiles. The dummy-linear specification is also somewhat below the Fourier series for most of the range between the 25th and 75th percentiles.

C. EOPP Data and Results

As a check on the NLSY functional form results, we now look at the evidence provided by EOPP. Unlike the NLSY, EOPP is not a longitudinal survey, and it only contains information on training at the start of the job, which causes difficulties for the rate of return analysis below. But EOPP does provide good measures of both formal and informal training. It also provides a measure of the number of weeks it takes a new employee to become fully qualified if he or she has the necessary school provided training but no experience in the job, which we refer to as "job complexity," as suggested by Barron, Berger, and Black (1999).[16]

EOPP's information on training comes from employers' reports about the number of hours the most recently hired worker spent in various training activities during his first three months of employment. We use the same measures of formal and informal training as previous papers using EOPP (see Barron, Black, and Loewenstein 1989 for example). Employers in EOPP provide information about the average wage paid to a worker who has been in the most recently filled position for two years, allowing one to estimate a pseudo fixed-effect equation. In the estimations that follow, the dependent variable is the difference between the log of the wage after two years and the log of the starting wage paid to the most recently hired worker. Besides the training variables, we include the following explanatory variables in all of our estimated equations: the most recently hired worker's age, gender, years of education, tenure, and dummy variables indicating whether the worked has received any vocational training or belongs to a union. In addition, we include the log of the number of employees at

16. For more information about the survey and the training questions, see Barron, Black, and Loewenstein (1989).

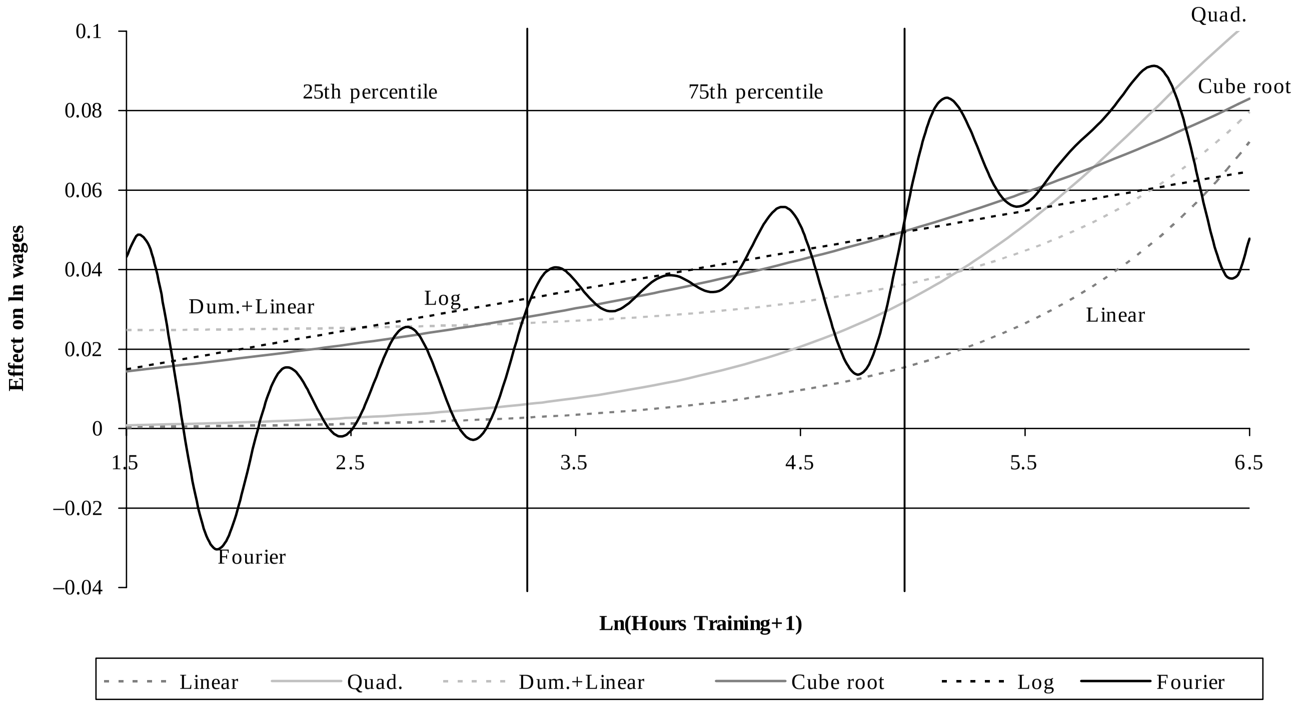

Figure 1
Predicted Effect of Training on Ln Wages, Various Specifications

the establishment, dummies indicating whether the most recently filled position was part-time or seasonal, two occupational dummies, and dummy variables for missing education, tenure, and union.[17] Finally, we also include as controls several variables that are less commonly found in other data sets—the most recently hired worker's relevant employment experience in jobs having some application to the position for which he was hired, relevant experience squared, and the log of the job complexity measure described above.

We exclude observations with missing values for any variables other than tenure, union, or years of education. We also exclude observations where wage growth is more than seven deviations above or below the sample mean. Finally, we exclude farm and government jobs. The resulting sample has 1,715 observations.

Sample means are reported in Table 3. Note that the bulk of training is informal. Ninety-five percent of workers receive informal training during the first three months of employment, but similar to the NLSY only 13 percent of workers receive formal training. And while mean informal training for those with any informal training is 132 hours, mean formal training for those with any formal training is only 72 hours. In a preliminary analysis we found that the log is the best-fitting simple specification for estimating the wage effect of informal training. Consequently, in our analysis of formal training, we include the log of informal training as one of our control variables. Our analysis of formal training is not sensitive to our treatment of informal training.

We again begin our analysis of the effect of formal training by estimating a Box-Cox model. The estimated value of λ is 0.376, which is quite close to the estimate from the NLSY. (Not surprisingly, since there are only 219 observations with positive formal training, the standard error for the EOPP estimate is quite high—0.291.)

The results for the various functional forms for training are shown in Table 4. The EOPP results are in general agreement with those from the NLSY. Once again, the linear specification performs the worst: like the NLSY data, the EOPP data indicate quite clearly that there are diminishing returns to training. The cube-root specification performs the best in the sample without outliers and second best in the complete sample. Furthermore, when one uses the cube-root specification, the estimated effect of training in the EOPP sample is similar to the estimated effect in the NLSY.

The quadratic specification fits best in the complete sample and comes closest to the estimated Fourier function, but in light of the fact that we only have 219 observations with positive formal training, we would not place much weight on the Fourier results. The simple cube-root and quadratic specifications both have a higher $\overline{R}^2$ than the Fourier series. The volatility of the quadratic specification between the EOPP and NLSY samples—both in fit and (as can be seen comparing Table 2 and Table 4) in size of the effect of training—makes us reluctant to recommend it as an alternative to the cube root.

17. Employers are (implicitly) asked about the starting wage paid to the most recently hired worker at the time he was hired, but about the average wage currently paid to workers with two years experience in the job. Since wages increase over time, tenure is positively correlated with wage growth.

Table 3
Descriptive Statistics, EOPP

Variable	Mean	Standard Deviation	Minimum	Maximum
Ln wage growth	0.19	0.20	–0.56	1.51
Formal training indicator	0.13	0.33	0.00	1.00
Informal training indicator	0.95	0.22	0.00	1.00
Hours of formal training, formal training > 0	72.27	101.14	1.00	640.00
Hours of informal training, informal training > 0	131.72	175.03	1.00	2,070.00
Ln (formal training + 1), formal training > 0	3.57	1.23	0.69	6.46
Ln (informal training + 1), informal training > 0	4.23	1.21	0.69	7.64
Ln number of weeks until fully trained	2.21	1.24	0.00	6.03
Years relevant experience	2.38	4.49	0.00	40.00
Relevant experience squared	25.76	108.31	0.00	1,600.00
Age	26.89	9.10	16.00	64.00
Years education	12.47	1.65	2.00	24.00
Vocational schooling	0.28	0.45	0.00	1.00
Temporary or seasonal job	0.15	0.36	0.00	1.00
Part-time job	0.21	0.41	0.00	1.00
Union	0.11	0.28	0.00	1.00
Ln establishment size	2.87	1.51	0.00	8.60
Female	0.45	0.50	0.00	1.00
Managerial/professional	0.11	0.31	0.00	1.00
Tenure	1.32	1.59	0.00	29.92
Other white-collar	0.57	0.50	0.00	1.00
Missing union	0.13	0.11	0.00	1.00
Missing years education	0.03	0.18	0.00	1.00
Missing tenure	0.03	0.18	0.00	1.00
Observation	1,715			

Table 4
Returns to Training for Different Functional Forms, EOPP

Specification	$\bar{R}^2$	Fraction Fourier Series Explained	Total Effect at Median
Complete sample			
No formal training variables	0.1756	—	—
Dummy	0.1807	0.587	0.044
Linear	0.1813	0.660	0.014
Quadratic	0.1837	0.951	0.031
Cube root	0.1834	0.873	0.047
Log	0.1830	0.840	0.052
Dummy + linear	0.1822	0.790	0.036
Fourier series	0.1833	—	0.041
Observations	1,715		
Training outliers omitted*			
No formal training variables	0.1749	—	—
Dummy	0.1801	0.574	0.045
Linear	0.1829	0.837	0.018
Quadratic	0.1831	0.963	0.028
Cube root	0.1834	0.899	0.050
Log	0.1827	0.847	0.053
Dummy + linear	0.1832	0.902	0.035
Fourier series	0.1830		0.041
Observations	1,713		

*Top 1 percent of formal training duration observations omitted.

III. Rates of Return to Training

Our best-fitting specifications in the NLSY indicate that 60 hours of formal training, the median positive amount of training, increases wages by 3–4 percent. The estimated effects of training in EOPP are even larger, as high as 5 percent for the median positive training of 38 hours. Relatively short training spells thus have substantial effects on wages. For comparison, current estimates of the effect of a year of school on wages are about 10 percent for the United States (see Jaeger 2003, for example). Here we examine how one can obtain estimates of the rate of return to the training investment from the coefficients in a wage regression. We take as our starting point a simple model in which a worker's wage always reflects his productivity. We then modify this model to take into account frictions in the wage-setting process.

A. *Rate of Return Calculations Using Coefficients on Lagged, Current, and Lead Training*

Consider a worker whose value of marginal product is given by $q = g(T)$, where T denotes training and where $g' \geq 0$. We allow workers with no training to have positive productivity—that is, $g(0) > 0$. But we assume that while receiving training, a worker produces nothing. For the time being, we neglect the direct cost of training. Assuming for expositional convenience that the job match is infinitely lived, the present value of the stream of output of an employee who receives T years of training from time t to $t + T$ is $V(T) = \int_{T+t}^{\infty} g(T)\exp(-r(\tau - t))\,d\tau = \frac{g(T)\exp(-rT)}{r}$, where r is the discount rate. The internal rate of return IRR for a training investment of T years is defined as the value of r such that $V(T) = V(0)$. Simple algebra establishes that $IRR = \frac{\ln(g(T)) - \ln(g(0))}{T}$.

It is easy to establish that the present value of output is maximized at a value of T where $MR(T) = r$ and $\frac{dMR(T)}{dT} < 0$, where $MR(T) \equiv \frac{d\ln(g(T))}{dT} = \frac{g'(T)}{g(T)}$. Given diminishing returns, infra-marginal returns will be greater than marginal returns and $IRR = (\int_{o}^{T} MR(u)du)(1/T) > MR(T)$. Thus, given rapidly diminishing returns to training, a high IRR clearly does not imply suboptimal investment in training. In contrast, assuming that our fixed-effect estimates of returns can be interpreted structurally, high marginal returns would imply suboptimal investment in training.

Note also that high observed average IRRs do not imply the existence of economic rents. If jobs with good training opportunities did offer economic rents, workers would enter these jobs, driving down output prices and wages. In equilibrium, the wage profile for a given job would depend upon the training it offers, but, other things the same, the present value of the wage stream would be equalized across all jobs.

How will the effect of training on productivity be reflected in wage growth? Suppose for the moment that the worker bears all the cost and realizes all the gain to training. Let w_t denote the worker's wage at time t and let T_t denote his accumulated training at time t. If wages were adjusted continuously, then we would have $w_t = g(T_t)$. That is, the wage at any moment in time is determined solely by the contemporaneous stock of completed training; lagged and lead values of training do not affect the wage.

In reality, frictions in wage setting prevent wages from being adjusted continuously. Consider an example where the worker is hired at time 0 and wages are adjusted once every year. Suppose that a training spell of length T starts at time τ_1 and ends at time τ_2, where $1 < \tau_1 < \tau_2 < 2$. Then the worker's wage is during the first year of employment ($0 \leq t < 1$). During the second year ($1 \leq t < 2$), the wage is given by: $w_t = \pi_1 g(0) + \pi_2 g(T)$, where $\pi_1 = \tau_1 - 1$ is the fraction of the second year that the worker works before receiving training and $\pi_2 = 2 - \tau_2$ is the fraction after the receipt of training. Wages during the period that training takes place are thus a weighted average of pre- and post-training productivity, with the weights adding up to less than one because there is no production during training itself. (If π_2 is sufficiently small relative to the time spent in training, $w_2 - w_1$ will be negative.) During the third year ($2 \leq t < 3$), the wage is given by $g(T)$. Note that the effect of training is spread over two periods—the period of training itself and the period after training has been completed.

Now consider a regression of wage observations on the stock of (completed) training accumulated on the job. Observations are recorded by a survey at times t_1, t_2, t_3, with $k - 1 < t_k < k$. To see how training should enter the regression, let us return to our example. Note that two cases are possible. If training is completed before the survey date t_2, then $T_1 = 0$ and $T_2 = T_3 = T$, so that *current* training causes the log wage at time t_2 to change from $\ln(g(0))$ to $\ln(\pi_1 g(0) + \pi_2 g(T))$. At time t_3, *lagged* training then causes an additional change in the log wage of the amount $\ln(g(T)) - \ln(\pi_1 g(0) + \pi_2 g(T))$. On the other hand, if training is completed after the survey date t_2, then $T_2 = T_1 = 0$ and $T_3 = T$. In this case, *lead* training causes the log wage at time t_2 to change from $\ln(g(0))$ to $\ln(\pi_1 g(0) + \pi_2 g(T))$. And at time t_3, *current* training causes an additional change in the log wage of the amount $\ln(g(T)) - \ln(\pi_1 g(0)+ \pi_2 g(T))$.

If the sample is a mixture of the two cases, then current, lagged, and lead training all belong in the wage equation. In our example, if the proportion p of individuals complete training T before the interview date t_2, the observed effect of lagged training is $p[\ln(g(T)) - \ln(\pi_1 g(0) + \pi_2 g(T))]$, the observed effect of current training is $(1 - p)[\ln(g(T)) - \ln(\pi_1 g(0) + \pi_2 g(T))] + p[\ln(\pi_1 g(0) + \pi_2 g(T)) - \ln(g(0))]$, and the observed effect of lead training is $(1 - p)[\ln(\pi_1 g(0) + \pi_2 g(T)) - \ln(g(0))]$. The total effect, $\ln(g(T)) - \ln(g(0))$, is the sum of these three effects. Accordingly, to estimate the IRR of training it is necessary to include one lead and one lag term.[18]

The foregoing has assumed that the worker bears all the costs (in terms of foregone production) and obtains all the returns to training. To the extent that the firm shares in the cost of training, the wage effect will underestimate the return to training in terms of productivity.[19] The observed wage effect is thus a lower bound (subject to caveats explained later).

B. Rate of Return Results

Table 5 shows results for the specifications considered in Table 2, with terms for lagged and lead training added. (Wage observations for the year 2000 were omitted, as lead training is not observed.) The functional form comparisons match those of Table 2, with observed wage effects about 25–35 percent higher. The order for K in the Fourier series expansion is 13 for the complete sample and 2 for the outlier-omitted sample; evidently the additional terms are needed only to track the behavior of the function for training outliers.

Setting a year equal to 2000 hours, we compute rates of return for T hours of training as $IRR(T) = 2000\sum_{-1}^{1} \beta_t f(T)/T$, where β_{-1}, β_0, and β_1 are the coefficients on lagged, current, and lead training. Rates of return at median positive training of 60 hours are in the 150–175 percent range for the better-fitting parsimonious specifica-

18. In our estimations, we set lagged training to zero the first period a worker is in a job. Lead training for the last period is set to the worker's final training in the job. If a worker leaves a job at time τ after survey date t_N but before survey date t_{N+1}, final training is obtained by adding training between t_N and τ to training at time t_N.

19. Other contracting situations are plausible. For example, the costs of training may be shared by workers who do not receive training. We leave these considerations for another paper.

Table 5
Rates of Return to Training for Different Functional Forms with Lagged and Lead Training, NLSY

Specification	$\bar{R}^2$	Fraction Fourier Series Explained	Total Effect at Median	Implied Rate of Return at Median (percent)
Complete sample				
No training variables	0.1949	—	—	—
Dummy	0.1961	0.580	0.042	140
Linear	0.1957	0.294	0.004	12
Quadratic	0.1959	0.359	0.007	23
Cube root	0.1969	0.752	0.045	149
Log	0.1968	0.745	0.053	175
Dummy + linear	0.1965	0.678	0.040	132
Fourier series	0.1982	—	0.059	197
Number of jobs	16,534			
Observations	69,800			
Training outliers omitted*				
No training variables	0.1939	—	—	—
Dummy	0.1951	0.746	0.042	140
Linear	0.1951	0.554	0.008	28
Quadratic	0.1956	0.777	0.018	59
Cube root	0.1959	0.959	0.048	159
Log	0.1958	0.942	0.053	178
Dummy + Linear	0.1957	0.906	0.040	134
Fourier series	0.1960	—	0.056	186
Number of jobs	16,502			
Observations	69,573			

*Top 1 percent of training duration omitted.

tions. Because series estimates potentially pick up local features of the wage-training function, the estimated return at a specific point may not be representative of returns over larger intervals and is likely to have a high standard error. Accordingly, for the Fourier series estimates, we calculate the mean return for the 25th through 75th percentiles of the distribution of positive training (to correspond to median training)—hereafter referred to as the mid-range return. The estimated mid-range return, shown in the first two rows of Table 6, is in the neighborhood of 180 percent.[20]

20. As expected, for the parametric estimators, the estimated mid-range returns are similar to the estimates at the median. The same turns out to be true for the Fourier series.

Table 6
Fourier Series Estimates of Mean Rates of Return for 25th–75th Percentiles of Positive Training Distribution (Standard Errors in Parentheses)

Without correction for heterogeneity in growth rates	
Complete sample	183
	(34)
Outlier-omitted sample*	178
	(33)
Corrected for heterogeneity in growth rates	
Complete sample	128
	(38)
Outlier-omitted sample*	124
	(36)
Corrected for heterogeneity in growth rates and promotions	
Complete sample	79
	(32)
Outlier-omitted sample*	75
	(31)
Corrected for heterogeneity in growth rates and promotions' effect on lead and current training coefficients	
Complete sample	88
	(33)
Outlier-omitted sample*	89
	(30)

*Top 1 percent of training duration omitted.

IV. Further Discussion and Interpretation of the Key Findings

Under the best-fitting specifications, the effect of formal training on wages is quite large. Rates of return for formal training estimated from the NLSY are in the 150–180 percent range for the median positive training of 60 hours. The effect of training on wages in EOPP is of comparable size. These numbers are much higher than, for example, estimated returns to schooling. The numbers also present a puzzle in view of the small proportion of jobs in both the NLSY and EOPP that have any formal training—13 percent in EOPP in the first three months, and 26 percent in the NLSY as of the last observation on the job. Taking the results literally, it appears that potentially profitable investments in training are not being made. In this section we discuss four potential explanations as to why estimated returns to training are so high: heterogeneity in wage growth, promotions, direct costs of training, and heterogeneity in returns to training.[21]

21. Measurement error is another potential explanation. A majority of our sample report receiving no formal training, and those who report positive formal training report varying amounts. Unlike the standard analysis, the mixed continuous-discrete nature of formal training means that measurement error may cause estimates of the effect of short spells to be biased upward. However, one can show that the bias in estimated returns at the geometric mean is likely to be small. See Frazis and Loewenstein (2003) for details.

A. Heterogeneity in Wage Growth

Our fixed-effect regressions control for factors whose effect on wages remains unchanged during a job match. However, unobserved factors that affect both wage *growth* and training will bias the fixed-effect estimates. To test whether individuals who receive more training tend to have higher wage growth even in the absence of training, we add interactions of tenure, tenure squared, and tenure cubed with the cube root of an individual's final observed training in the current job to the regression. (A preliminary analysis showed that the cube root of final training fit best.) If workers with higher wage growth self-select into training, then the estimated effect of final training on wage growth should be positive and the other training coefficients should fall.[22]

This is in fact what we observe. The third and fourth rows of Table 6 shows mid-range rates of return for the Fourier series.[23] The rate of return to training falls by 55 percentage points to about 125 percent, and the final training interactions are jointly significant at the 1 percent level.[24] The interaction coefficients imply that respondents who end up with 60 hours of training average about 0.8 percent per year more rapid wage growth initially and about 0.6 percent per year after 2.5 years.

The rate of return for the Fourier series is higher than that for the cube-root specification both at the median itself and between the 25th and 75th percentiles. Table 7 shows the coefficients for the leads and lags in the cube-root specification, with and without final training interactions. The lag and lead coefficients decline greatly in size.[25]

B. Promotions

While general heterogeneity in wage growth does not completely explain the large estimated returns to training, it is possible that employees are offered training after increases in their job responsibilities. This might cause us to falsely attribute wage increases to training that are in fact due to promotions. Both the NLSY and EOPP contain data on promotions, so we can estimate the extent that correcting for promotions reduces our estimates of the effect of training.

The 1988–90 NLSY surveys asked respondents whether their job responsibilities had increased since the last interview. Respondents also were asked whether they had

22. Pischke (2001) controls for wage growth heterogeneity by including an interaction between tenure and an individual fixed effect in a regression that already includes a (noninteracted) fixed-effect term. Our approach allows for a flexible tenure interaction—for example, individuals who acquire more training may tend to have higher wage growth in the first few years of tenure but not later.

23. As in the previous specification, $K = 13$ for the complete sample and $K = 2$ for the outlier-omitted sample.

24. A previous draft of this paper found no effect of the final training-tenure interactions on returns to training. The difference in results is due to the addition of tenure cubed to the specification and improved measurement of final training.

25. Note that the final training-tenure interaction coefficients are identified because interactions of stocks of (current, lead, and lagged) training with tenure are excluded from our regression, as are leads and lags beyond one period. Allowing a limited set of training-tenure interactions provides some evidence that the returns to training decline with tenure over the first few years on the job. The estimated return to training at the sample median tenure is similar to the text. See Frazis and Loewenstein (2003) for details.

Table 7
Selected Coefficients and Rates of Return, Cube Root Specification

	Full Sample		Outliers Omitted*	
Lead training$^{1/3}$	0.0024	0.0006	0.0027	0.0010
	(0.0016)	(0.0016)	(0.0018)	(0.0018)
Current training$^{1/3}$	0.0051	0.0043	0.0050	0.0044
	(0.0014)	(0.0014)	(0.0015)	(0.0015)
Lagged training$^{1/3}$	0.0040	0.0013	0.0045	0.0019
	(0.0011)	(0.0012)	(0.0012)	(0.0013)
Final training$^{1/3}$ × tenure		0.0021		0.0021
		(0.0008)		(0.0009)
Final training$^{1/3}$ × tenure^2		–0.0001		–0.0001
		(0.0001)		(0.0001)
Final training$^{1/3}$ × tenure^3/100		0.0001		0.0002
		(0.0005)		(0.0005)
Effect of training at median positive hours	0.0448	0.0246	0.0476	0.0285
	(0.0070)	(0.0084)	(0.0075)	(0.0088)
Rate of return to training at median positive hours	149	82	159	95

*Top 1 percent of training duration omitted.

received a promotion and, if promoted, whether responsibilities had increased as a result of the promotion. In 1996–2000, respondents were asked separate questions about changes in job responsibilities and promotions. We focus on changes in job responsibilities because a "promotion" after training may merely be a recognition of the worker's increased productivity. Using promotion variables produces similar results.

We total changes in responsibilities within each job separately over the years 1987–90 and 1994–2000 and estimate a wage equation over both subperiods (where the job is unchanged, separate fixed-effects are estimated in each subperiod).[26] We find that adding this variable to a Fourier series specification that includes final training interactions reduces the mid-range rate of return by 48 percentage points in the outlier-omitted sample, implying a rate of return of 75 percent as reported in Table 6. Similarly, in the cube-root specification, adding the change in responsibilities variable reduces the sum of the training coefficients by 0.0030 and the estimated rate of return by about 39 percentage points, producing a rate of return of 56 percent in the outlier-omitted sample.

It is very likely that there is mutual causation between training and promotions. For example, in the *SEPT95* sample of employees (Frazis et al. 1997), of those who

26. To create a uniform variable about changes in job responsibilities, in the 1988-90 period we count an individual as experiencing a change in responsibilities if (a) he answers affirmatively to the change in responsibilities question or (b) indicates that his responsibilities have changed as a result of a promotion.

received formal training from their current employer, 14 percent reported receiving a promotion when training was satisfactorily completed and 40 percent reported that training was necessary for future advancement (categories are not mutually exclusive). Thus, not surprisingly, training helps workers get subsequent promotions. We have an identification problem; while giving an able worker more duties may increase productivity in the absence of training, a worker's improved ability to carry out more advanced job duties should properly be considered part of the return to training. The above specification attributes all promotion-induced wage growth to promotions per se as opposed to the training that may have made the promotions possible. The estimated 40–50 percentage point reduction in the effect of training is clearly too large.

A more reasonable way of accounting for promotions is to control for them when estimating the effect of the current stock of training, but not when estimating the effect of lagged training—that is, to calculate the return to training as $\beta_{Lt} + \beta_{Lt+1} + \beta_{St-1}$, where β_{Lt} is the training coefficient for period t in the long regression including promotions and β_{St} is the coefficient for period t in the short regression omitting promotions. This procedure attributes promotion-induced wage growth to the promotion if the promotion occurs roughly concurrently with or some time before training, and to training if the promotion is realized some time after training. This approach, which is still probably too conservative, yields a reduction in the estimated rate of return from promotions of 34 (30) percentage points using the Fourier series estimates (cube-root specification), resulting in a rate of return of 89 (64) percent.

In contrast to the NLSY, the EOPP data provide no indication that the estimated return to training is partly due to the effect of promotions. Employers in EOPP are asked whether the last worker hired has received a promotion and, if so, how many months after being hired. When one includes in the wage growth equation a dummy variable indicating whether a worker has received a promotion within two years of being hired, one obtains a promotion coefficient that is positive and significant, but there is virtually no effect on the formal training coefficient.

C. Direct Costs of Training

The 1995 Survey of Employer Provided Training (SEPT95) estimated that, in its sampling frame of firms with 50 or more employees, salaries of trainers and other surveyed components of the direct cost of training totaled \$300 per employee in 1994. The survey also estimated that wages paid to employees while in formal training totaled \$224 over the period May-October 1995 (Frazis et al. 1997). Pro-rating the wage cost of employees to a full year, the wages paid to workers receiving training appear to account for only about 60 percent of the total costs of training; other direct costs account for the remaining 40 percent. Applying this to our previous results, we obtain an estimated rate of return of about 40–50 percent.

D. Heterogeneity in Returns to Training

One strongly suspects that our estimated returns are greater than could be realized by workers without formal training were they to get such training. Because jobs differ in the skills they require, it makes sense that the returns to training vary across jobs. The NLSY provides direct evidence of heterogeneity in returns.

We interact the cube root of the current stock of training with job characteristics in the NLSY: the two occupational dummies and the part-time indicator. (The specification also includes final training interactions.) Results are shown in Table 8. Both the managerial and professional dummy and the part-time dummy have strong positive effects on the returns to training, with managerial and professional jobs having an 80 percentage points greater rate of return to training of 60 hours than do blue-collar jobs (the difference is strongly significant). Managerial and professional employees are more likely to receive formal training, which is consistent with their higher returns. The positive effect of part-time is harder to interpret since part-time status is negatively associated with training. One possibility is that it reflects higher

Table 8
Selected Coefficients and Rates of Return, Cube Root Specification with Job Characteristics Interactions, NLSY, Outlier-omitted Sample

	Coefficient	
Lead training$^{1/3}$	0.0010	
	(0.0018)	
Current training$^{1/3}$	0.0012	
	(0.0017)	
Lagged training$^{1/3}$	0.0016	
	(0.0013)	
Initial occupation, managerial/professional × current training$^{1/3}$	0.0063	
	(0.0015)	
Initial occupation, other white collar × current training$^{1/3}$	0.0015	
	(0.0013)	
Part-time × current training$^{1/3}$	0.0133	
	(0.0049)	
	Effect at 60 Hours	Rate of Return at 60 Hours
Blue collar	0.0152	51
	(0.0091)	(30)
Managerial/professional	0.0397	133
	(0.0099)	(33)
Other white-collar	0.0212	71
	(0.0094)	(31)
Part-time blue collar	0.0672	112*
	(0.0205)	(34)

*Calculated at work-year of 1,000 hours.

required effects of training on productivity to make investing in a part-time employee worthwhile.[27]

These results are evidence that the return to training varies greatly across jobs.[28] If some of the heterogeneity in returns is unobservable, as seems likely, then our results do not reflect the returns to training that could be obtained by the average member of the population. To see this, consider the following wage model that abstracts from covariates other than training:

$$(2)\quad \ln W_{it} = \alpha_i + \beta_i \varphi(T_{it}) + e_{it},$$

where $E(e_{it}) = E(\alpha_i) = 0$, $E(\beta_i) = \bar{\beta}$, and e_{it} is independent of α and β.

Both α and β are potentially correlated with T. There is ample evidence that training is higher for more productive workers, presumably because their cost of training is lower and/or their return to training is higher.[29] A correlation between unmeasured ability α and training causes OLS estimates of the return to training to be upward biased. Fixed-effect estimation eliminates this bias, but it does not purge the effect of a correlation between β and T.

For example, consider a situation where we have two periods of data, with training always equal to zero when $t = 1$ and varying across the sample when $t = 2$. The expected value of the return to training estimated by fixed effects (which, in this case, is equivalent to first differences) is given by:

$$\begin{aligned}(3)\quad f(T_0) &= E\left(\ln W_{i2} \middle| T_{i2} = T_0\right) - E\left(\ln W_{i1} \middle| T_{i2} = T_0\right)\\ &= E\left(\alpha_i \middle| T_{i2} = T_0\right) - E\left(\alpha_i \middle| T_{i2} = T_0\right) + E\left(\beta_i \varphi(T_0) \middle| T_{i2} = T_0\right)\\ &= E\left(\beta_i \middle| T_{i2} = T_0\right) \varphi\left(T_0\right).\end{aligned}$$

One can distinguish between the return to training for the average member of the population and the return to training for the trained (see Heckman and Robb 1985). Fixed-effect regressions do not estimate the return to training for the average member of the population $\bar{\beta}\,\varphi\,(T_0)$, but, as is clear from Equation 3, consistently estimate the effect of a given amount of training *for those with that amount of training*.[30] In particular, our high estimated returns to short spells of training are not overestimates of

27. While training part-time blue-collar workers for 60 hours raises their wage by 6.7 percent, the implied rate of return is slightly less than the return to training full-time managers and professionals if one calculates the part-time rate using 1,000 hours for a work year instead of 2,000 hours.

28. EOPP also contains evidence that returns to training vary across jobs. Specifically, the returns to an aggregate of formal and informal training (with weights chosen to fit our wage data) are higher for more complex jobs. See Frazis and Loewenstein (2003) for details.

29. For example, see Barron, Berger, and Black (1999).

30. Note that the example given, with zero training in the first period followed by varying amounts in the second period, is exactly the situation in EOPP. The situation is more complicated in the multiperiod NLSY data set, where the estimated return $g(T_0)$ will partly reflect average returns and partly reflect marginal returns. When we omit observations with (within-job) accumulated training greater than zero but less than final observed training—thus bringing the situation closer to that in EOPP—the results are virtually identical to those in Table 2.

the return to training for those with such spells. However, this does not mean that one would expect individuals who do not receive formal training to realize such returns if they were trained. Indeed, any reasonable model would predict that $E(\beta_i|T = T_0) > E(\beta_i|T = 0)$: trained individuals should tend to have a higher return than those with no training.

Without the appropriate structural restrictions, it is not possible to estimate the expected return to training of workers who do not receive training. Similar comments apply to estimates of the marginal return to training, which will be estimated as

$$(4) \quad f'(T_0) = E(\beta_i|T = T_0)\,\varphi'(T_0) + \frac{\partial E(\beta_i|T = T_0)}{\partial T}\,\varphi(T_0),$$

and which will exceed $E(\beta_i|T = T_0)\varphi'(T_0)$ if $\frac{\partial E(\beta_i|T = T_0)}{\partial T} > 0$: estimation of φ' is confounded by a composition effect stemming from the fact that individuals with more training can be expected to have a higher return.

E. Summary

Heterogeneity in wage growth, promotions, and direct costs are all partial explanations for the high estimated rates of return to training appearing in Tables 5 and 6. After correcting for these factors, we are left with annualized returns in the neighborhood of 40–50 percent for 60 hours.[31] These returns are very likely an underestimate in that they do not reflect cost-sharing with the employer. Heterogeneity in returns is one explanation as to how returns to formal training can be so high while most workers do not get training. Although those with formal training of 60 hours do have annualized returns to training of at least 40–50 percent, these returns cannot be extrapolated to the untrained.

V. Conclusion

This paper has investigated the related questions of the functional form and magnitude of the wage returns to formal training. Our results from both the NLSY and EOPP indicate that the return to an extra hour of training diminishes sharply with the amount of training received. A cube root specification generally fits the data best, but the log specification also does well. The linear specification always fits the data poorly and substantially understates the effect of training, and the quadratic specification is volatile.

Our wage-setting model indicates that to estimate the true return to training, the wage regression must include not just the current stock of training, but also lead and lagged training. Our best-fitting specifications indicate that there are very substantial returns to the initial interval of formal training, 150–180 percent for the median positive value. Taking into account heterogeneity in wage growth, promotions, and the

31. This estimate turns out to be similar to those obtained by Bartel and Mincer, but this is by coincidence as our estimate is obtained very differently. Unlike Bartel and Mincer, we use a nonlinear specification, allow for one period lead and lagged effects as implied by theory, and control for heterogeneity in wage growth, promotions, and the direct cost of training.

direct cost of training reduces these estimates to the neighborhood of 40–50 percent. This estimated return is an underestimate since it does not take into account cost-sharing with the employer. We found evidence of heterogeneity in returns and concluded that our estimates reflected the average return to training for the trained and could not be extrapolated to the untrained.

While a fair amount of research on the econometrics of heterogeneous returns has recently been published (for example, Angrist, Imbens, and Rubin 1996; Heckman 1997), structural estimation of returns to training when there is heterogeneity presents challenges. Not only is it difficult to find a plausible instrument, but the mixed continuous-discrete structure complicates the problem. We leave a more complete analysis of heterogeneity in returns to training as a topic for future research.

References

Andrews, Donald W. K. 1991. "Asymptotic Optimality of Generalized C_L, Cross-validation, and Generalized Cross-validation in Regression with Heteroskedastic Errors." *Journal of Econometrics* 47(2):359–77.

Angrist, Joshua, Guido Imbens, and Donald Rubin. 1996. "Identification of Causal Effects using Instrumental Variables." *Journal of the American Statistical Association,* 91(434):444–72.

Barron, John M., Mark C. Berger, and Dan A. Black. 1997. *On-the-Job Training*. Kalamazoo Mich.: W. E. Upjohn Institute of Employment Research.

———. 1999. "Do Workers Pay for On-the-Job Training?" *Journal of Human Resources* 34(2):235–52.

Barron, John M., Dan A. Black, and Mark A. Loewenstein. 1989. "Job Matching and On-the-Job Training." *Journal of Labor Economics* 7(1):1–19.

Bartel, Ann P. 1995. "Training, Wage Growth, and Job Performance: Evidence from a Company Database." *Journal of Labor Economics* 13(3):401–25.

Frazis, Harley, Maury Gittleman, Michael Horrigan, and Mary Joyce. 1997. "Formal and Informal Training: Evidence from a Matched Employee-Employer Survey." In *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, v. 9, ed. Gary Libecap, 47–82. Greenwich, Conn.: JAI Press.

Frazis, Harley, and Mark A. Loewenstein. 2003. "Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation." BLS Working Paper 367. Washington: Bureau of Labor Statistics.

Gallant, A. Ronald. 1981. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form." *Journal of Econometrics* 15(2):211–45.

Greene, William H. 2000. *Econometric Analysis*, 4th ed. Upper Saddle River, New Jersey: Prentice-Hall.

Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32(3):441–62.

Heckman, James, and Solomon Polachek. 1974. "Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship." *Journal of the American Statistical Association* 69(346):350–54.

Heckman, James, and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data,* eds. James Heckman and Burton Singer, 156–245. New York: Wiley.

Jaeger, David A. 2003. "Estimating the Returns to Education Using the Newest Current Population Survey Education Questions." *Economics Letters* 78(3): 385–94.

Lengermann, Paul A. 1999. “How Long Do the Benefits of Training Last? Evidence of Long Term Effects Across Current and Previous Employers.” *Research in Labor Economics* 18:439–61.

Lillard, Lee A., and Hong W. Tan. 1986. *Training: Who Gets It and What Are Its Effects on Employment and Earnings?* Santa Monica California: RAND Corporation.

Loewenstein, Mark A., and James R. Spletzer. 1996. “Belated Training: The Relationship Between Training, Tenure, and Wages.” Bureau of Labor Statistics. Unpublished

Lynch, Lisa M. 1992. “Private Sector Training and the Earnings of Young Workers.” *American Economic Review* 82(1):299–312.

Mincer, Jacob. 1989. “Human Capital and the Labor Market: A Review of Current Research.” *Educational Researcher* 18(4):27–34.

Parent, Daniel. 1999. “Wages and Mobility: The Impact of Employer-Provided Training.” *Journal of Labor Economics* 17(2):298–317.

Pischke, Jörn-Steffen. 2001. “Continuous Training in Germany.” *Journal of Population Economics* 14(3):523–48.