BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF EMPLOYMENT RESEARCH
AND PROGRAM DEVELOPMENT

Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation

Harley Frazis and Mark A. Loewenstein

Working Paper 325
August 1999

The views expressed are those of the author(s) and do not reflect the policies of the U.S. Bureau of Labor Statistics or the views of other staff members. The authors thank Dan Black, Marilyn Manser, Terra McKinnish, and participants in a session of the 1998 Southern Economic Association meetings for helpful comments. We thank Alexander Eidelman for assistance in constructing the NLSY dataset, Hector Rodriguez for help with figures, and James Spletzer both for help with the NLSY data and for comments.

Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation

Harley Frazis
Bureau of Labor Statistics
2 Massachusetts Ave. NE
Suite 4945
Washington D.C. 20212

Phone: 202-606-7394
Frazis_H@bls.gov

Mark A. Loewenstein
Bureau of Labor Statistics
2 Massachusetts Ave. NE
Suite 4130
Washington D.C. 20212

Phone: 202-606-7385
Loewenstein_M@bls.gov

The authors thank Dan Black, Marilyn Manser, Terra McKinnish, and participants in a session of the 1998 Southern Economic Association meetings for helpful comments. We thank Alexander Eidelman for assistance in constructing the NLSY dataset, 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.

Abstract

This paper examines the appropriate functional form and the size of the wage returns to training. In both the National Longitudinal Survey of Youth (NLSY) and Employer Opportunity Pilot Project (EOPP) datasets a log specification fits best. In the NLSY, the full effect of training occurs with a lag as long as two years, training on previous jobs is a substitute for training on the current job, and the return to training declines with labor market experience. The EOPP data indicate that formal and informal training are perfect substitutes; however, an hour of formal training has a much greater effect on wages than does an hour of informal training.

We find very large returns to formal training in both the NLSY and EOPP. The mixed continuous-discrete nature of the training variable means that measurement error can cause estimates of the effects of short spells of training to be biased upward, but we demonstrate that the maximum upward bias in estimated returns at the geometric mean is minimal. Heterogeneity in returns is a more plausible explanation of the high estimated return to training; in the EOPP data, the return to training is significantly higher in more complex jobs. With unobserved heterogeneity in returns, our estimates can 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 datasets with direct measures of training. Studies find support for the human capital model's prediction that a worker's wage is positively related to past investments in his training.[1] Indeed, Brown (1989) reports that "within-firm wage growth is mainly determined by contemporaneous productivity growth". Similarly, Barron, Black, and Loewenstein (1989) note that "training is one of only a few variables affecting wage and productivity growth."

While there is a widespread agreement that training is positively associated with wage growth, relatively little attention has been paid to the size of the effect.[2] Researchers have paid even less attention to the choice of the appropriate functional form. Indeed, the variation in functional forms across studies makes it difficult to compare estimated rates of return. This difficulty is compounded by the fact that researchers using different functional forms have tended to use different datasets: while users of the Employer Opportunity Pilot Project (EOPP) data and the closely related Small Business

---

[1] A non-exhaustive list of references here includes Altonji and Spletzer (1991), Barron, Berger, and Black (1999), Barron, Black and Loewenstein (1989, 1993), Bartel (1995), Brown (1989), Lengermann (1999), Lillard and Tan (1986), Loewenstein and Spletzer (1996, 1998, 1999a), Lynch (1992), Mincer (1988), Pischke (1999) and Veum (1995).

[2] We are aware of two attempts to calculate rates of return to on-the-job training. 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 dataset, 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% 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.

Administration (SBA) data have generally used log specifications (for example, see Barron, Black, and Loewenstein 1989 and Barron, Black, and Berger 1999), researchers using the National Longitudinal Study of Youth (NLSY) have used linear specifications (for example, Lynch 1992, Parent 1999) or specifications that estimate the return to a spell of training without making use of information on the duration of the spell (Loewenstein and Spletzer 1996, Lengermann 1999).

In this paper, we use the NLSY and EOPP datasets to narrow the range of plausible estimates of the return to on-the-job training by carefully investigating the choice of the appropriate functional form. The NLSY and EOPP datasets complement each other very well. The NLSY is a longitudinal survey of workers and provides the best information on formal training at all levels of tenure. EOPP is an employer survey and provides the best information on formal and informal training at the start of the job. Thus, for example, the NLSY allows one to examine the time pattern of the return to training and the effect of training in previous jobs on the return to training in the current job, while the EOPP data allow one to examine the extent to which informal training is a substitute for formal training.

Our key findings from the NLSY and EOPP are very similar. The results from both datasets indicate that the return to an extra hour of training diminishes sharply with the amount of training received. In fact, the pattern of returns is accounted for very well by a log specification. Our estimates from both datasets imply there are very high returns to the initial interval of formal training, especially for workers with low levels of tenure and experience. For example, the NLSY data show a wage return to the first 40 hours of formal training as large as 8 percent for persons with low levels of tenure and experience (in the EOPP data, this return is 6 percent), which is roughly the same magnitude as estimated returns to a *year* of schooling.

The large size of the estimated returns to formal training raises the question as to exactly what these estimates represent. As explained below, because of the mixed continuous-discrete nature of formal training, measurement error can potentially lead to the overestimation of the returns to short spells. We model the effect of measurement error and find that under reasonable assumptions, the maximum upward bias in estimated returns at the geometric mean of training is minimal.

Heterogeneity is a more plausible explanation of the high estimated return to training. Indeed, the EOPP dataset provides direct evidence of heterogeneity in returns. As discussed below, with unobserved heterogeneity in returns, our estimates can be regarded as the (average) return to training for the trained. But despite our control for heterogeneity in wage levels by means of a fixed effect, this return cannot be extrapolated to the untrained.

It is interesting to contrast our results with the literature on the returns to schooling. Some have suggested that the effects of measurement error on the one hand, and ability bias and heterogeneity in returns on the other, roughly cancel each other out, leaving OLS estimates as a good guide to the return to schooling for the average member of the population (Ashenfelter and Zimmerman 1997, Ashenfelter and Rouse 1998). This is clearly not the case for formal training—taking heterogeneity into account appears essential to making sense of estimated returns.

The remainder of the paper is organized as follows. We analyze the NLSY data in section II and the EOPP data in section III. We then turn to questions of interpretation in section IV, and conclude in section V.

II. Analysis of the NLSY

<u>Data</u>

The NLSY is a dataset of 12,686 individuals who were aged 14 to 21 in 1979. These youth have been interviewed annually since 1979, and the response rate has been 90 percent or greater in each year. We use data from the 1979 through 1994 surveys.[3] The training section of the survey begins with the question, "Since [the date of the 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.[4]

---

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

[4] Individuals are not asked about the number of weeks spent in training if their training spell is in progress at the time of the interview, but instead this information is obtained in a subsequent interview after the training spell is completed. Thus, besides being asked about training spells that began since the last interview, individuals are also asked about spells that were still in progress at the time of the last interview.

The training questions were changed somewhat in 1988. From 1979-1986, detailed information was obtained only on training spells that lasted longer than one month.[5] 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. Besides conditioning on the fact that a spell lasts less than one month, our imputations also condition on an individual's age.[6]

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.[7] 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 classification.

The two key training variables used in the empirical work to follow are total

---

[5] Training questions were not asked in 1987.

[6] In the later surveys, individuals were explicitly asked about both the weekly duration of training and the year and month that a training spell began and ended. In the early surveys, individuals were asked about the year and month that a training spell began and ended, but were not explicitly asked about the number of weeks that a training spell lasted. Inspection of the post-1987 data reveals that 4 weeks is the best estimate for the weekly duration of training when a training spell ends in the subsequent month, 8 weeks is the best estimate for the weekly duration of training when a training spell ends in the second month after it began, and so on. Individuals are not asked about the starting year and month of a training spell that was in progress at the time of the last interview. For the pre-1987 data, we have obtained this information by carrying it forward from the year in which a training spell initially began. We have not had to do this for the post-1987 data because in the later years individuals were explicitly asked about the weekly duration of training spells that were in progress at the time of the last interview.

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

accumulated completed training on the current job and total accumulated completed training on past jobs. These variables are obtained by adding the training a worker has completed in the current year to the training he has received in all previous years.

Basic Results

Our basic specification is:

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

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 training on the current job, f(·) 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, homoscedastic and uncorrelated with X_{ijt} and across jobs, persons and periods. All specifications are run as fixed-effect regressions within jobs. The vector X includes experience and experience squared, tenure and tenure squared, age, a dummy for ever married, years of education (which occasionally changes within a job), year dummies, and interactions of tenure with the following: age, female, AFQT,[8] years of education, ever married, union, two dummies for initial occupation in the job, and missing AFQT and missing union. As additional controls for training, we include a count of spells with missing training duration (most of these occur before 1988) and a dummy for training ongoing at the time of the interview.

We exclude observations with missing values on variables other than AFQT and union. We also exclude observations with real wages below $1 or above $100 in 1982-84

[8] Specifically, the residual from a regression of AFQT on dummies for year of birth.

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). We also exclude the military subsample, active members of the armed forces, the self-employed, those in farm occupations, and observations where the respondent was enrolled in school at any time between interviews. The resulting sample has 61,033 observations from 15,876 jobs. Descriptive statistics are shown in table 1.

The results for different functional forms for training are shown in table 2. The top panel shows results using the entire sample. We include an "incidence" specification that counts total number of training spells. The "log" specification is ln(T+1), where T is number of hours of training.[9] The table shows adjusted R^2s (explained variance as a proportion of within-job variance) and total effects of training at the median number of hours of training for those with positive training. The differences in fit appear slight. However, the best-fitting specification—the log—increases adjusted R^2 several times as much as the worst-fitting specification—the linear—relative to the fit excluding training variables. The quadratic specification and the incidence specification are little improvement on the linear, while the square root specification is close to the log. The quadratic specification has the advantage of allowing one to check for decreasing returns since the estimated coefficient on the quadratic term provides a test of the null hypothesis

---

[9] Note that this is a special case of the more general specification $f(T)=\ln(sT+1)$. Non-linear least squares yields an estimated value of s that equals .18, but is not significantly different from one. Note also that the model $E(y|\ X,\ T) = a_0+X\beta + \gamma f(T)$, $f(T)= \ln(sT+c)$, where c is a parameter to be estimated, is not identified, so setting c equal to 1 is without loss of generality.

of constant or increasing returns to training. The null is rejected at the 5 percent level ($t=-1.95$).

The implied effect of the median hours of training (or the median number of spells) differs by more than a factor of 30 between the different specifications, with the log specification showing the largest effect and the linear and quadratic specifications apparently greatly understating the effect. The implied annualized rate of return of training, assuming a work-year of 2000 hours, ranges from 3 percent to 92 percent.

One might suspect that the better fit of the log and square 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, and also the top one percent of the distribution of number of spells. The total effect of the median amount of positive training is similar in the sample without outliers to the total sample for all specifications with the possible exception of the quadratic. The log specification is still the best fitting; there is no marked improvement in the fit of the other specifications relative to the log. The t statistic on the quadratic term increases in magnitude to –3.79, providing a stronger rejection of the hypothesis of constant returns.

Our best fitting specification, the log, implies large initial returns to training that decline steeply. 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 estimate two more general functional forms. The first of these is a spline specification with transition points at the 25^{th}, 50^{th}, and 75^{th} percentiles of the positive training distribution.

Spline results for the entire sample are shown in table 3. The return to an hour of training declines steeply from .00078 between the 0^{th} and 25^{th} percentiles to .00022 between the 50^{th} and 75^{th} percentiles to .00001 between the 75^{th} and 100^{th} percentiles. The negative return between the 25^{th} and 50^{th} percentile is not statistically significant and most likely due to sampling error. The results for the sample with outliers omitted are similar, although in this case the return between the 75^{th} and 100^{th} percentiles is negative but not statistically significant.

As a more flexible functional form, we estimate a Fourier series expansion (Gallant, 1981). A Fourier series expansion of K terms of a function f(T) is:

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

In practice, linear and quadratic terms are usually added. Moreover, for non-periodic functions the variable T needs to be transformed to a variable T* such that $0 < T^* < 2\pi$. The expansion is then implemented as:

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

The Fourier expansion has the property that the differences between both the level of the true function and its derivatives, and the level of the Fourier expansion and its derivatives, 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).[10]

In our case, due to the essentially log-normal distribution of training,[11] it is more convenient to work with the log of training as a basis for the Fourier expansion rather than training itself. Our transformation is $T^*=0.001+\ln(T+1)/2$, which has a range of .001 to 4.24 in our outlier omitted data. We added terms to the expansion until $\bar{R}^2$ decreases; this turned out to be K=4.

We use the statistic $Q^2 \equiv 1 - \frac{\sum (f(T) - f^*(T^*))^2}{\sum (f^*(T^*) - 0)^2}$ as a convenient summary measure of the closeness of fit between an arbitrary specification f(T) and the estimated Fourier series f*(T*).[12] 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 series and f(T) relative to a specification which omits training. As can be seen in the third column of table 2, the log specification is closest to the Fourier series, and the linear specification is the furthest. Indeed, the log specification explains about 90 percent of the squared distance between

[10] Other semi-parametric estimation methods are harder to adapt to the fixed-effect setup. Li and Stengos (1996) consider fixed-effect estimation of β_1, but it is not possible to estimate f(T) directly using their method.

[11] We estimated the Box-Cox transformation to normality $(T^\lambda - 1)/\lambda$, where λ=0 corresponds to a log-normal distribution and λ=1 to a normal distribution. For the positive training sample, λ is estimated to be -.03.

[12] We are grateful to Dan Black for suggesting this type of statistic.

the Fourier series and zero, while the linear specification only explains about 30 percent.

Figure 1 plots the estimated effect of training for the linear, log, quadratic, square root, and Fourier series specifications in the sample without outliers. The effect is plotted against log training since a linear scale would overly compress the range where the data are concentrated. As indicated by the Q^2 statistic, the figure shows that the log specification fits the basic pattern of returns in the Fourier series expansion better than the other functional forms over most of the range of the data.[13] (Not surprisingly, the Fourier series is fairly erratic, especially in ranges of the data near the maximum or minimum.) One can see that the log specification is closer to the Fourier series than the other specifications in an interval starting below the 25th and ending above the 75th percentile.

Why do the functional forms other than the log track the Fourier series so poorly, even in the middle of the positive training distribution? In our fixed-effect regressions, observations with large deviations of training from average training will have a disproportionately large effect on the training coefficient.[14] (Indeed, the justification for discarding training outliers stems from the fact that erroneous observations in the tails will have particularly damaging effects.) Specifications such as the linear should tend to predict better in the right tail of the distribution and worse in the middle of the training

---

[13] As indicated in table 2, the relative performance of the log specification is even better when we do not drop outliers from the sample.

[14] Specifically, the coefficient on training is given by $\hat{\beta}_2 = \frac{\sum (f(T) - \hat{f}(T))(\ln W - \hat{\omega})}{\sum (f(T) - \hat{f}(T))^2}$, where $\hat{f}(T)$ and $\hat{\omega}$ denote the predicted values of f(T) and ln W from regressions of f(T) and ln W on X and the fixed effects. Note that $\hat{b}_2$ is a weighted sum of the $(\ln W - \hat{\omega})$ observations, with the absolute value of the weights proportional to the absolute value of $f(T) - \hat{f}(T)$.

distribution than specifications like the log that compress the training distribution. The linear function's tendency to fit the right tail will lead to an especially poor fit in the middle of the training distribution when linearity is a misspecification.

As a final specification test, we added squared and cubed terms (i.e., $\ln(T)^2$ and $\ln(T)^3$) to the log specification. These terms were not significant (p=.35). We accordingly adopt the log specification for the remainder of this section.

The Timing of the Returns to Training

The above specifications implicitly assume that all of the wage returns to training occur shortly after training is completed. We now examine specifications with lagged training to determine the time pattern of the return to training. Table 4 shows returns to log training for training lagged up to three years, with lagged training coded as zero if the respondent was not with the current employer during the relevant period. As can be seen, lagged training has sizable and statistically significant effects on wages, but the lag does not extend beyond two years. Failure to take into account lagged effects results in a sizable understatement of the total returns to training. Summing the training coefficients for the specification with training lagged two years (or three years), the total return to a log hour of training is .012, compared to .008 without lags. At the median hours of training, this corresponds to an annualized return of 137 percent.

Substitutability with Other Forms of On-the-Job Training.

One might expect training with a previous employer to be a substitute for training with the current employer. More broadly, one might expect tenure on the current job, or labor market experience in general, to be a substitute for formal training. Here we examine interactions of log training on the current job with other forms of on-the-job human capital accumulation.

In column 1 of table 5, we interact (log) training on the current job (lagged twice) with log training on previous jobs.[15] We sum coefficients over all lags to obtain both the total wage effect of training on the current job and the interaction effect of training on previous jobs. We find a small but statistically significant negative interaction, implying that previous job training and current job training are substitutes. The effect on the coefficients for current job training is relatively slight. The implied effect of training for a respondent with no previous job training is about 10 percent higher than was found in table 4.

We interact training on the current job with tenure and tenure squared in column 2. There is a substantial interaction effect. The effect of a one-unit increase in log training declines at the rate of .002 log points per year of tenure at zero tenure (calculated as the sum of the coefficients on the interactions of linear tenure with log training on the current job); a one-unit increase in log training now increases log wages by .022 at zero tenure and zero training on previous jobs. In column 3, we further interact training on the current job with experience and experience squared. We find that training is a substitute for early career experience. The interaction with (linear) tenure is cut in half and is no longer significant, while the effect of a one-unit increase in log training declines at the rate of .002 log points per year of experience at zero experience.[16] The increase in log wages from a one-unit increase in log training with zero experience (and hence zero tenure) and no previous job training is estimated as .029, more than twice the return to log training estimated in table 4, although it should be noted that there are few observations with zero experience.[17,18]

To make clearer the effect of the various forms of human capital acquisition on the return to training, table 6 shows estimated returns to 40 hours of training for 12 hypothetical respondents with various values of the relevant characteristics. We assign either zero or the median positive value (78.9) of training on the current job, and similarly for training on previous jobs (here the median positive value is 200 hours). We assign the 25th, 50th, and 75th percentile values of tenure and experience, varying both together (that is, we assign a respondent with the 25th percentile value of experience the 25th percentile value of tenure).

As one might expect, training on the current job is the primary determinant of the return to further training. The return to training declines by a factor of nine as one moves

---

[15] Note that the noninteracted duration of training at previous employers is absorbed into the fixed effect. Work by Loewenstein and Spletzer (1998,1999b) and Lengermann (1999) indicates that training received at a previous employer results in a higher wage at the current employer.

[16] The declines in the return to training with tenure and experience are non-linear. In results not shown, interacting a linear tenure term with training without also interacting tenure squared shows little effect of tenure on the return to training, and similarly for experience.

[17] While observations where the respondent is enrolled in school are excluded from the wage regression, we do not exclude experience gained while enrolled in school from our experience measure.

[18] The change in the NLSY training sequence does not appear to have influenced our estimation results. In several of the above specifications, including the specification in column 3 of table 5, we have interacted the training variables with an indicator for year before 1987. These interactions are never either singly or jointly statistically significant at conventional levels.

from zero to the median positive value; the annualized rates of return to 40 hours of training for those with no training on the current job are in the 200-400 percent range, compared to 20 to 45 percent for those with the median positive value. The effects of training in previous jobs, tenure, and experience on the return to further training are smaller, but still substantial. As one simultaneously goes from the 25th to the 75th percentile of tenure and experience and from zero to the median positive value of training on previous jobs, the return to training is cut by more than half.

III. Analysis of the EOPP

Data

The NLSY provides strong evidence that returns to formal training decline greatly with the quantity of such training. While the NLSY has the important advantage of being a longitudinal survey, making it possible to keep track of the training that individuals receive over their working life, the NLSY does not contain information on informal training that is useful for our present purposes.[19] To look at the nature of the return to informal training, we now turn to the Employer Opportunity Pilot Project (EOPP) survey. Unlike the NLSY, EOPP is not a longitudinal survey, and it only contains information on

[19] The NLSY began asking detailed questions about informal on-the-job training in the 1993 survey. After completing the formal training questions, an individual in the NLSY is asked whether he had to learn new job skills in the past 12 months because of some change at work. The ensuing sequence of "informal training" questions in the 1993 NLSY is designed to measure training that was not already recorded in the preceding sequence of formal training questions. However, as noted by Loewenstein and Spletzer (1999a), the routing patterns in the informal training questions limit their use as a source of information regarding on-the-job skill acquisition: individuals who did not experience changes at work within the past 12 months are not routed into the detailed training questions. The informal training questions have been revamped in the most recent survey.

training at the start of the job. However, EOPP has the advantage of providing good measures of both formal and informal training.[20] Information on formal training comes from employers' reports about the number of hours specially trained personnel spent giving formal training to the most recently hired worker during his first three months of employment. We obtain a measure of informal training by summing (1) the number of hours that line supervisors and management personnel spent giving the most recently hired worker informal individualized training and extra supervision, (2) the number of hours that co-workers spent away from other tasks in providing the most recently hired worker with informal individualized training, and (3) the number of hours that a new worker typically spends watching others do the job rather than do it himself. Our previous analysis of the NLSY indicates that the pattern of wage returns to training is accounted for very well by a log specification. We would like to know whether the same is true for informal training. An additional question of interest is the precise nature of the substitution between informal and formal training.

EOPP is not a longitudinal study, so it is not possible to estimate a true fixed-effect wage equation. However, 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 logarithm of the wage after two years and the logarithm of the starting wage paid to the most recently hired worker.

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

Besides the training variables, we include the following explanatory variables in all of our estimated equations: the most recently hired worker's age, number of years of education, gender, a dummy variable indicating whether or not the most recently hired worker belonged to a union, the logarithm of the number of employees at the establishment, and two occupational dummies. Also included as control variables are several variables that are less commonly found in other datasets–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 logarithm of the number of weeks it takes a new employee in the most recently filled position to become fully trained and qualified if he or she has the necessary school provided training but no experience in the job.

We exclude observations with missing values for any of the variables. The resulting sample has 1,550 observations. Sample means are reported in table 7. Clearly, 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 134 hours, mean formal training for those with any formal training is only 75 hours. Thus, not only are workers much more likely to receive informal training than formal training, but informal spells last longer.

Results

Following the previous literature (Barron, Black, and Loewenstein 1989, for example), we first look at specifications in which formal and informal training are simply summed to obtain total training. Note that this specification makes two implicit assumptions. First, it presumes that formal and informal training are perfect substitutes.

Second, it presumes that the rate of substitution is one-to-one since an hour of informal training has exactly the same effect on wages as does an hour of formal training. An equivalent way of expressing the latter idea is that formal and informal training have equal weights in constructing the aggregate “total training”. In our first set of equations, log wage growth is regressed against total training and other control variables. The results for different functional forms are shown in table 8. Table 8 mirrors table 2 and shows adjusted R^2s and total effects of training at the median number of hours of training. The top panel shows results using the entire sample, while the bottom panel reports the estimation results when one omits observations in the top 1 percent of the training distribution. As with the NLSY equations, the differences in fit are slight since all specifications capture the tendency for wages to increase with the stock of training. Nevertheless, the EOPP results, like the NLSY results, provide strong indication of diminishing returns to training; the linear specification is clearly the worst-fitting.

The estimated return to training in EOPP is remarkably close to the estimated return in the NLSY. Recall (from column 2 of table 5) that the total return (allowing for lags as long as two years) to a log hour of formal training when tenure is zero is .022 in the NLSY. If one does not control for informal training in EOPP, one obtains a coefficient of .018 on log hours of formal training.[21] (As reported in column 2 of table 9, this coefficient falls to .016 when one adds log hours of informal training to the equation.)

Unlike the case with the NLSY, the choice of the best-fitting functional form is not so clear in EOPP: the square root specification yields a higher adjusted R^2 than the log

[21] Loewenstein and Spletzer (1999a), using a slightly different EOPP sample, report a coefficient of .021 on log hours of formal training when informal training is not controlled for.

specification, which in turn yields a higher adjusted R^2 than the quadratic. This order is reversed when one omits training outliers. Before attempting to pass judgment on the appropriate functional form, we examine more carefully the assumptions that formal and informal training are perfect substitutes and that an hour of formal training is equivalent to an hour of informal training, assumptions that are implicit in all of the specifications in table 8. The most natural place to begin investigating this question is with the log specification.

For convenience, the specification using the logarithm of total training is reported again in column 1 of table 9. Column 2 of table 9 shows what happens when one includes the logarithm of formal training and informal training as separate arguments. Note that the second estimation yields a better fit, as measured by the adjusted R^2. As can be seen in column 3, adding an interaction term between formal and informal training does not yield a further improvement in fit.

The equations in columns 1 and 2 can be viewed as special cases of a more general specification of the form,

$$\text{(2a)} \quad \ln(W_{it}) - \ln(W_{i0}) = \ln(\psi(T_{formal},T_{informal})) + X_i\beta + \varepsilon_{it},$$

$$\text{(2b)} \quad \psi(T_{formal},T_{informal}) = A(b(T_{formal}+1)^{-\rho} + (1\text{-}b)\,(T_{informal}+1)^{-\rho})^{-k/\rho},$$

where W_{i0} and W_{it} are worker i's starting wage and post-training wage, respectively, and T_{formal} and $T_{informal}$ are the quantities of formal and informal training. Note that the function $\psi(\cdot,\cdot)$ is a generalized constant elasticity of substitution production function.[22] The

---

[22]As with the simple log specification, we have modified the standard "constant elasticity of substitution" specification to ensure that the wage function is defined when formal and informal training are both zero. In the standard constant returns to scale specification, $k = 1$. Our more general specification allows for

parameter ρ is closely related to the elasticity of substitution, σ: $\sigma = 1/(1+\rho)$. When ρ is positive and large, σ is small and informal training cannot be easily substituted for formal training. In the limit when $\rho \to 0$, the production function becomes generalized Cobb-Douglas, or $\psi(T_{formal},T_{informal}) = A((T_{formal})^{b}(T_{informal})^{1-b})^{k}$. Finally, when $\rho \to -1$, isoquants become linear. In this case, formal and informal training are perfect substitutes and can be combined to form a single aggregate; the parameter b indicates the relative weights that should be placed on formal and informal training when they are aggregated into total training. The two types of training are weighted equally only when b = .5.

We estimated equation (2) with nonlinear least squares. The estimation results are summarized in column 4 of table 9. If we do not restrict the parameter ρ to the economically sensible range, the unconstrained nonlinear least squares estimate of ρ is – 2.97; this parameter is estimated very imprecisely, with a standard error of 9.38. A value of ρ less than – 1 can be ruled out on economic grounds since it implies that isoquants are not convex to the origin. When one imposes the corner restriction that $\rho = -1$, the adjusted R^2 increases to .0952: a test of the restriction based on the difference in the sum of squared residuals yields a chi-squared statistic with a p-value of .541.[23] The estimate of b in the restricted specification is .95. The EOPP data thus indicate that formal and informal training are very good substitutes--indeed, our constrained point estimate is that they are perfect substitutes--and that an hour of formal training should be counted as being

---

increasing (k > 1) and decreasing returns (k < 1). As is true with the NLSY data, the EOPP data strongly imply diminishing returns.

worth 19 hours of informal training.[24] These results are not driven by outliers. When one omits observations in the top 1 percent of the training distribution, the estimated values of ρ and b are both unchanged.

We estimated linear, quadratic, and square root specifications treating formal and informal training as perfect substitutes and compared these specifications with ones analogous to columns 2 and 3 of table 9. For each functional form, the best fitting specification is that in which formal and informal training are treated as perfect substitutes. As reported in table 10, the estimated weight on an hour of formal training is always much higher than that on an hour of informal training. As was the case with the NLSY, the log specification does somewhat better than the square root specification, which has an adjusted R^2 of 0.0930. The log specification also does significantly better than the quadratic.

Unlike the NLSY, the estimated return to training in EOPP does not appear to be affected by experience. When one interacts relevant experience and relevant experience squared with aggregate training, there is little change in the coefficient on aggregate training and the interaction terms are insignificant.

The various specifications are graphed in figures 2 and 3 along with the estimated fourth order Fourier series (as with the NLSY, the length of the expansion that maximizes adjusted R^2; interestingly, when one does not omit outliers, the log specification actually

---

[23] Imposing the restriction that $\rho = -1$, 2b) reduces to $\psi(T_{formal}, T_{informal})=A\,(bT_{formal} + (1-b)T_{informal} + 1)^k$. This is a special case of the more general specification $\psi(T_{formal}, T_{informal})=A\,(s\,(bT_{formal} + (1-b)T_{informal}) + 1)^k$. The non-linear least squares point estimate of s equals 1.28, but is very imprecise.

[24] Note, however, that the estimated wage elasticity with respect to formal training (.26) is only a little higher than the estimated wage elasticity with respect to informal training (.15).

has a higher adjusted R^2 than the Fourier series). The weights assigned to formal and informal training are allowed to vary among all the functional forms. Figure 2 portrays the estimated return to formal training when informal training is assigned its median value for the subsample that receives positive formal training. Similarly, figure 3 depicts the estimated return to informal training for those with the median value of formal training (zero). When one confines one's attention to the ranges of the graphs where most of the data are located, the linear specification is again clearly the worst-fitting while the log appears to be the best-fitting. The square root and quadratic specifications perform similarly for formal training. In the case of informal training, the square root does about as well as the log and considerably better than the quadratic.

In summary, the EOPP data, like the NLSY data, indicate that the return to training diminishes sharply with training. However, a comparison of figures 1, 2 and 3 reveals that the differences in fit between the linear and diminishing return specifications are not as great as in the NLSY. The same conclusion emerges when one examines the increase in adjusted R^2 relative to the fit excluding training variables and when one looks at the proportion of the squared distance between the Fourier series and zero explained by the various specifications (reported in column 3 of table 10). In the EOPP data, the effect of the median hours of formal training on wages is higher by a factor of three for the log specification than for the linear. This is a substantial difference, but smaller than the factor of 30 that we obtained from the NLSY data.

As noted above, the EOPP data indicate that formal and informal training are sufficiently close substitutes that they can meaningfully be summed to form a measure of aggregate training. In forming this sum, formal training receives a much higher weight than informal training, implying that an hour of formal training has a significantly greater wage effect than an hour of informal training. When one calculates the estimated effect of median training on a worker's wage, one finds that the lower marginal effect of informal training is in large part offset by the fact that workers receive much more informal training than formal training. Specifically, adopting the log specification, the median hours of formal training increase a worker's wage by 6.2 percent, while the median hours of informal training raise the wage by 3.8 percent. The implied annualized rates of return of training are 308 percent and 103 percent, respectively.

IV. Further Discussion and Interpretation of the Key Findings

Under the best fitting specification, the estimated returns to formal training in the NLSY are very large. The results in table 6 show returns to the first 40 hours of formal training as large as 8 percent for persons with low levels of tenure and experience, roughly the same magnitude as estimated returns to a *year* of schooling. Marginal returns at the median amount of formal training are an order of magnitude smaller but still quite large; the smallest number in table 6 still implies an annualized rate of return that exceeds 20 percent. Similarly in the EOPP data, the estimated high weight of formal training in our training aggregate implies a very high return to formal training. The first 40 hours of formal training, for a worker with the median hours of informal training, increase wages by 6.2 percent.

Only a minority of respondents in both the NLSY and EOPP have any formal training. This is puzzling in view of the high estimated returns. Estimated wage gains should be, if anything, less than gains in productivity.[25] Taking the results literally, it would appear that potentially profitable investments in training are not being made. One possibility is that the high returns for short spells are an artifact of measurement error; another is that they reflect heterogeneity in returns to training. We discuss each in turn.

Measurement Error

Substantial measurement error in training has been reported by Barron, Berger, and Black (1997a). In the standard analysis, measurement error results in estimates that are biased downward. However, the case of formal training is more complicated because of its mixed continuous-discrete nature: a majority of our sample report receiving no formal training, and those who report positive formal training report varying amounts. This mixed continuous-discrete structure implies that, as explained below, estimates of the effect of short spells of training may well be biased upward.

To determine the likely effects of measurement error on our OLS results, let T^* denote true training and T denote observed training. In addition, let $g(T_0)$ denote the return to training for those whose true training is T_0. Abstracting from other covariates for convenience, $g(T_0) = E(\ln W|T^*=T_0) - E(\ln W|T^*=0)$, where presumably $g' > 0$ and

[25] In the classic theory of human capital (Becker 1975), workers incur all the costs and realize all the returns to general human capital, but share the costs and returns to specific human capital with employers. Wage growth should therefore equal productivity growth when training is general but be less than productivity growth when training is specific. Loewenstein and Spletzer (1998) argue and provide evidence that contract enforcement considerations can lead to employers sharing the returns and costs to even purely general training. Also see Acemoglu and Pischke (1999). Using subjective measures of productivity in the EOPP and SBA data, Barron, Berger, and Black (1997b) find that the productivity growth associated with training is several times the wage growth.

$g'' \leq 0$. Since we do not observe true training, our reduced form estimation does not yield the function g, but instead yields f, where $f(T_0) = E(\ln W|T=T_0) - E(\ln W|T=0)$ is the expected return to training for an individual whose observed training is T_0. (We assume throughout that we consistently estimate f; we do not consider distortions caused by incorrect functional forms.)

One can distinguish between two types of measurement error: misclassification of training and error in the duration of spells that are classified correctly. Misclassification in turn can be subdivided into forgotten training, where $T = 0$ but $T^* > 0$, and false training, where $T^*=0$ but $T > 0$. Provided that both types of misclassification error are independent of the residual ε in equation (1), misclassification unambiguously reduces the observed return to training, $f(T_0)$. To see this, note that if there is any forgotten training, $E(\ln W|T=0) > E(\ln W|T^*=0)$. And if there is any false training of length T_0, then $E(\ln W|T= T_0) < E(\ln W|T= T_0, T^* > 0)$. The greater is either type of misclassification error, the smaller is the observed return to training, $f(T_0)$.

To gain intuition on the effects of duration error, consider figure 4. For ease of exposition, the figure assumes away misclassification. Line G in the figure represents the true function $g(T)$, which goes through the origin. Under standard conditions, measurement error in the positive training sample will flatten the observed function, as shown in line F in the figure. If there is no misclassification error, $E(\ln W|T=0) = E(\ln W|T^*=0)$, so earnings of those with no training will be consistently estimated. However, for any level of training $0 < T_0 < M$, $E(\ln W|T= T_0) > E(\ln W|T^*= T_0)$, implying that the returns to training in this range will be overestimated.

To formalize this intuition and to bound the effects of measurement error, we need to put further structure on the problem. For convenience, we disregard classification error. We model duration error x as multiplicative, so that $T=T^*x$ for $T^* > 0$. Let $u = \ln(x) = \ln(T) - \ln(T^*)$ denote the measurement error in (non-misclassified) log training. We assume that u is distributed independently of $\ln(T^*)$ and that $E(u) = 0$. Letting $n(\cdot)$ denote the density for $\ln(T^*)$ and $f(\cdot)$ denote the density for log measurement error u, the conditional density for $\ln(T^*)$ given observed (non-misclassified) training $T= T_0$ is given by

$$(3) \qquad \eta(\ln(T^*), T_0) = \frac{f(\ln(T_0) - \ln(T^*))n(\ln(T^*))}{\int_0^\infty f(\ln(\mathrm{T}_0) - x)n(x)dx} \; .$$

The densities $n(\cdot)$ and $f(\cdot)$ are assumed to be unimodal and symmetric. We should note that our measurement error assumptions are consistent with our data in that the training distributions in both the NLSY and EOPP are both approximately log-normal.[26] In addition, in the NLSY reported training hours for each spell are the product of reported hours per week and reported spell duration in weeks, strongly implying a multiplicative element to the measurement error.

Let $h(\cdot)$ be the function implicitly defined by $h(\ln(T)) = g(T)$. Taking a second-order Taylor expansion of $h(\cdot)$ around $\ln(T_0)$, the expected return to training for an

---

[26] Estimating the Box-Cox transformation to normality $(\mathrm{T}^{\lambda} - 1)/\lambda$ yields an estimate of λ of .03 for the EOPP positive formal training sample. Recall that λ=0 corresponds to log-normality, and that our estimate of λ is –.03 in the NLSY. Quantile plots also show that log-normality is a good approximation in both datasets.

individual with observed non-misclassified training $T = T_0$ can be expressed as

$$\text{(4)} \qquad \mathrm{E}(\ln W|T = T_0) = g(T_0) + h'(\ln(T_0))\mathrm{E}(\ln(T^*|T = T_0) - \ln(T_0))$$

$$+ (1/2)\, h''(\ln(T_0))\mathrm{E}((\ln(T^*) - \ln(T_0))^2|T = T_0) + R(T_0),$$

where $R(T_0)$ is the error in the approximation.[27]

Note that when observed training T_0 is low, measurement error is more likely to be negative than positive, which means that the expected value of true training will exceed T_0. Conversely, the expected value of true training will be less than observed training if observed training is high. In fact, let $m^* \equiv \exp(\mathrm{E}(\ln(T^*)))$ denote the geometric mean of the training distribution. Then under our assumption that the densities $n(\cdot)$ and $f(\cdot)$ are unimodal and symmetric, one can show that

$$\text{(5)} \qquad \mathrm{E}(\ln(T^*)|T = T_0) \overset{<}{\underset{>}{=}} \ln(T_0) \text{ as } T_0 \overset{>}{\underset{<}{=}} m^* \;.$$

Consequently, for observed training at the geometric mean, (4) reduces to

$$\text{(6)} \qquad \mathrm{E}(\ln W|T = m^*) = g(m^*) + (1/2)\, h''(\ln(m^*))\, s^2_{\ln(T^*)|T=m^*} + R(m^*),$$

where $s^2_{\ln(T^*)|T=m^*}$ is the conditional variance of the log of true training when $T = m^*$.

---

[27] Note that

(a) $\mathrm{E}(\ln W|T = T_0) = \int_0^\infty h(\ln(x))\eta(\ln(x), T_0)\,dx$.

Taking a second-order Taylor expansion around $\ln(T_0)$, one can write

(b) $h(\ln(x)) = h(\ln(T_0)) + h'(\ln(T_0))(\ln(x) - \ln(T_0)) + (1/2)h''(\ln(T_0))(\ln(x) - \ln(T_0))^2 + r(x)$,

where the term $r(x)$ denotes an error term. Substituting (b) into (a) yields (4), where $R(T_0) = \mathrm{E}(r(x)|T=T_0)$ $\equiv \int_0^\infty r(x)\eta(\ln(x), T_0)\,dx.$

If the return to training $g(\cdot)$ is linear in log training, then the function $h(\cdot)$ is linear and $h''(T_0) = R(T_0) = 0$. It thus follows immediately from (6) that $E(\ln W|T = m^*) = g(m^*)$, i.e., the estimated return to training at the geometric mean should equal the true return. In the NLSY data, the implied annualized rate of return at the geometric mean of 88 hours is 85 percent for our simplest log specification, and 126 percent when one allows for lags. In the EOPP data, the annualized rate of return at the geometric mean of 37 hours is an even higher 330 percent. (The higher rate of return in EOPP can be attributed to the fact that the workers in the survey are just starting their current employment relationship and consequently have lower accumulated training on the current job.[28]) Clearly, measurement error cannot explain the high estimated return to training at the geometric mean if the return to training is log-linear.

If the true return to training declines at a greater than logarithmic rate, then $h(\cdot)$ is concave and $(1/2)\, h''(m^*)\, s^{2}_{\ln(T^*)|T=m^*} + R(m^*) < 0$. In this case, it follows from (6) that $E(\ln W|T= m^*) < g(m^*)$: the estimated return to observed training at (or not too far below) the geometric mean will be less than the true return. Thus, measurement error can also not explain the high estimated return to training if the true return to training declines at a greater than logarithmic rate.

Finally, if the true return to training declines at a slower than logarithmic rate, then $h(\cdot)$ is convex and the estimated return to non-misclassified observed training at (or not too far above) the geometric mean will exceed the true return. To estimate the potential

[28] Decreasing returns implies that the rate of return is inversely related to spell length. As discussed above, the return to training is inversely related to job tenure.

upward bias, let the return to true training be given by $g(T^*) = c(T^*)^a$, $0 < a < 1$, which means that $h(\mathrm{x}) = c(\exp(x))^a$ and

$$(7) \qquad h''(m^*) = \alpha^2 h(m^*).$$

In addition, suppose that measurement error and true training are both distributed lognormally. This implies that observed training is also distributed lognormally, which as noted above is a good approximation with our data. Given lognormality, one can show that[29]

$$(8) \qquad s^2_{\ln(T^*)|T=m^*} \leq (1/4)s^2_{\ln(T)}\,.$$

Furthermore, calculations reveal that the proportional error term, $R(m^*)/\,g(m^*)$ is negligible. It therefore follows from (6) and (8) that the maximum possible proportional upward bias in the estimated return to training at the geometric mean is $(1/8)s^2_{\ln(T)}\ h''(\ln(m^*))/h(\ln(m^*)) = (1/8)s^2_{\ln(T)}\ a^2$. The variance of observed log training in the NLSY is 2.41. In EOPP, the variance of observed log formal training is 1.54. Thus, if $a = 1/2$, the proportional bias in the estimated return to geometric mean training is less than 7.5% in the NLSY and 4.9% in EOPP. Even if the return to training is linear in

---

[29] Letting $s^2_{\ln(T^*)}$, $s^2_{\ln(T)}$, and s^2_u denote the unconditional variances of ln(T*), ln(T), and u, the coefficient of linear correlation between ln(T*) and ln(T) is given by

$$r = \frac{\mathrm{cov}(\ln(T^*),\ln(T))}{s_{\ln(T^*)}s_{\ln(T)}} = \frac{\mathrm{E}(\ln(T^*)\ln(T)) - \mathrm{E}(\ln(T^*))\mathrm{E}(\ln(T))}{s_{\ln(T^*)}s_{\ln(T)}} = \frac{\mathrm{E}(\ln(T^*)^2) - \mathrm{E}(\ln(T^*))^2}{s_{\ln(T^*)}s_{\ln(T)}} = \frac{s_{\ln(T^*)}}{\sqrt{s^2_{\ln(T^*)} + s^2_{u}}}\,.$$

If ln(T*) and ln(T) are normal, the conditional variance of ln(T*) is given by

$$\text{(a)} \quad \mathrm{var}(\ln(T^*)|\ln(T_0)) = (1\text{-}r^2)\,s^2_{\ln(T^*)} = \frac{s^2_u s^2_{\ln(T^*)}}{s^2_u + s^2_{\ln(T^*)}} = \frac{s^2_u(s^2_{\ln(T)} - s^2_u)}{s^2_{\ln(T)}}\,.$$

Partially differentiating (a) with respect to s^2_u, one finds that holding $s^2_{\ln(T)}$ constant, var(ln(T*)|ln(T_0))

training, i.e., *a* = 1, the proportional bias in the estimated return to geometric mean training is about 30% in the NLSY and 20% in EOPP. As discussed above, misclassification error will reduce these amounts. Thus, even if the true return to training declines at a slower than logarithmic rate, measurement error does not appear to explain the high estimated returns to training

If one drops that part of the sample that reports not having received training and estimates the returns to training only for the subsample reporting positive training, our measurement error problem reduces to the standard one and the estimated return to training should be biased downward. When we estimate returns to training in the NLSY using only the subsample reporting positive training, the log functional form is still the best fitting (but the square root is close) and the returns to training increase slightly. Similar results obtain in EOPP. These results need to be interpreted cautiously because the parameter estimates are not very precise, but they suggest that either misclassification is important in tempering the upward bias for short spells in the full sample or that measurement error may not be such a serious problem after all.[30]

---

is maximized when $s_u^2 = (1/2)\, s_{\ln(T)}^2$. Substituting into (a), one obtains the result that var(ln(T*)|ln(T_0)) $\leq (1/4) s_{\ln(T)}^2$.

[30] Our discussion neglects one potential complication in the NLSY: the fact that our measure of training stocks is not derived from a single questionnaire item, but is the sum of training flows accumulated across periods, each component of which is subject to misclassification and duration error. This would imply, among other things, that duration error in the stocks of training in the positive training sample would include the effects of misclassification error in the flows, as within the same job some spells of training are forgotten and some activities are misclassified as training, described above as "false training". Accounting for this would complicate our analysis considerably. The EOPP, with a single formal training item, is not subject to this problem. We conclude from the similarity of the results between the two datasets that this complication is probably not a major concern.

In conclusion, even if one allows for the maximum possible measurement error bias, the estimated return to training is still very high. Either the required rate of return to training investments is very high, or our results cannot be taken as a guide to the return to formal training for the untrained.

Heterogeneity

There are reasons why investments in training may have a high required rate of return. If the skills involved are firm-specific, the time horizon over which the returns to training are realized is the worker's tenure with the firm. This may be much shorter than the working life over which investments in schooling are realized.[31] Formal training may also have a high direct cost, especially if specialized skills are involved. However, we think it improbable that these factors completely explain our results.

If a week of formal training raises productivity by 8 percent, a worker need only be with a firm for 13 weeks for a training investment to break even. We would expect direct costs to be a major reason for a low training incidence in small firms, which cannot take advantage of economies of scale in training. However, while the incidence of training increases with establishment size, in our NLSY data slightly less than 50 percent of respondents from establishments with over 1,000 employees ever receive any formal training during the time they are in our wage sample, and similarly for respondents from establishments with over 2,000 or over 5,000 employees.

One strongly suspects that our estimated returns are greater than could be realized by workers without formal training were they to get such training. Since the skills required for different jobs are heterogeneous, it makes sense that the returns to training differ across jobs. The EOPP dataset provides direct evidence of heterogeneity in returns.

Recall that one of the control variables in our wage growth regression is the number of weeks it takes a new employee in the most recently filled position to become fully trained and qualified if he or she has the necessary school provided training but no experience in the job. Barron, Berger, and Black (1999) suggest that this variable provides a measure of job complexity. Consistent with this interpretation, the log number of weeks until fully qualified is positively related to wage growth. The last column in table 9 reports the effects of interacting hours of aggregate training with the number of weeks it takes a new employee without previous experience to become fully qualified. The coefficient on the interacted variable is positive and quite large. And adding the interacted variable to the wage growth equation reduces the coefficients on non-interacted aggregate hours of training and non-interacted weeks until fully qualified (not reported in the table) by factors of three and nine, respectively; in fact, the coefficients on both non-interacted variables are no longer significantly different from zero. These results provide strong evidence that the return to training varies greatly across jobs.

In addition to heterogeneity in returns to training, heterogeneity in wage growth may also affect our results. Unobserved factors that affect both wage growth and training will bias fixed-effect estimates of the return to training. To test whether individuals who receive more training tend to have higher wage growth even in the absence of training, we have modified the NLSY wage equation to include interactions of tenure and tenure

---

[31] Interestingly, Loewenstein and Spletzer (1999b) provide evidence that a substantial portion of on-the-job training is general. Nevertheless, Royalty (1996) does find empirical support for the proposition that

squared with an individual's final observed training in the current job. If workers with higher wage growth self-select into training, then the coefficient on the tenure - terminal training interaction term should be positive and the coefficients on current and lagged training should fall.[32] However, this turns out not to be the case. When added to the specification in column 3 of table 4, the tenure- terminal training interaction terms are not significant and have little effect on the key training coefficients. We conclude that for our data, heterogeneity in wage growth has little effect on the estimated return to training; heterogeneity in returns to training appears to be far more important.

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. This is in spite of our control for heterogeneity in wage levels by means of the fixed effect. To see this, consider the follow simplified wage model,

(9) $\ln W_{it} = a_i + \beta_i g(T_{it}) + e_{it}$,

where $E(e_{it}) = E(a_i) = 0$, $E(\beta_i) = \bar{\beta}$, and e_{it} is independent of α and β. For convenience, we now disregard measurement error and again abstract from other covariates in the wage equation.

Both α and β are potentially correlated with T. There is ample evidence that indicates that training is higher for more productive workers,[33] presumably because their cost of training is lower and/or their return to training is higher. If the cost of training is

---

training is less frequent on high-turnover jobs.

[32] Note the similarity between our use of terminal training and Abraham and Farber's (1987) use of completed tenure to control for unobserved heterogeneity in estimating the effect of tenure on wages.

[33] For example, see Barron, Berger, and Black (1999).

lower for more able individuals in more productive jobs, that is, if cov(α_i, T) > 0, then OLS estimates of the return to training will be biased upward.

Fixed-effect estimation eliminates any potential bias stemming from a positive correlation between unmeasured ability α and training. However, fixed-effect estimates of the return to training do not purge the effect of a correlation between β and *T*. The EOPP data provide evidence of just such a correlation. We noted above that the return to training is higher for individuals who are in jobs that require more time to be fully qualified. Loewenstein and Spletzer (1999) demonstrate that hours of aggregate training are strongly positively correlated with the number of weeks it takes a worker without experience to be fully qualified.

To analyze the bias in fixed-effect estimation, consider a situation where we have two periods of data, with training always equal to 0 when t=1 and varying across the sample when t=2. The expected value of the return to training estimated by fixed effects (first differences) is given by:

$$(10) \quad f(T_0) = E(\ln W_{i2}|T_{i2}=T_0) - E(\ln W_{i1}|T_{i2}=T_0)$$

$$= E(\alpha_i|T_{i2}= T_0) - E(\alpha_i|T_{i2}= T_0) + E(\beta_i g(T_0)| T_{i2}= T_0)$$

$$= E(\beta_i| T_{i2}= T_0)g(T_0).$$

One can distinguish between the return to training for the average member of the population and the return to training for the trained (see, for example, Heckman and Robb 1985 and Heckman 1997). Fixed-effect regressions do not estimate the return to training for the average member of the population $\bar{b}\, g(T_0)$, but, as is clear from (10), consistently estimate the effect of a given amount of training *for those with that amount of training*.[34] In particular, our high estimated returns to short spells of training are not overestimates of 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 have realized such returns had they been trained. Indeed, any reasonable model would predict that $E(\beta_i|T=T_0) > E(\beta_i|T=0)$: individuals with training 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

$$f'(T_0) = E(b_i|T=T_0)g'(T_0) + \frac{\partial E(b_i \mid T = T_0)}{\partial T} g(T_0), \qquad (11)$$

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

To summarize the effects of measurement error and heterogeneity in returns to training on estimated returns: Under reasonable assumptions, average returns at or above the geometric mean of training are underestimated or at worst slightly overestimated due

[34] 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. As with measurement error (see fn. 30), the situation is more complicated in the multiperiod NLSY dataset, 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.

to measurement error. Heterogeneity in returns does not affect this conclusion as long as estimated returns to a given amount of training are interpreted as the average return for those with that amount of training. Measurement error in duration will cause marginal returns to training to be underestimated,[35] while heterogeneity in returns will likely cause marginal returns to be overestimated (a situation similar to that found in the literature on returns to schooling; see, for example, Ashenfelter and Rouse 1998). We conclude that those with training around the geometric mean of the NLSY (EOPP) do have annualized returns to training of at least 100 (300) percent, but that these returns cannot be extrapolated to the untrained.

V. Conclusion

This paper has investigated the related questions of the functional form of the wage returns to training and their magnitude. We find that the logarithmic form appears to fit best. In the EOPP dataset, which contains extensive data on both formal and informal training, we find that formal and informal training are perfect substitutes, but an hour of formal training has a much larger effect on wages than an hour of informal training.

In both the NLSY and EOPP datasets, we find very large returns to formal training when we use the best-fitting functional form. These rates of return are an order of magnitude higher than those implied by the frequently-used linear specification,

[35] False training may cause marginal returns to be overestimated in ranges where the frequency of false training is declining. We regard this as likely to be of small importance in practice.

showing that choice of functional form has a large effect. Given returns of this magnitude, it is puzzling that only a minority of each sample receives formal training. It is very unlikely that the high return to training can be due to measurement error. Heterogeneity of returns is a more compelling explanation of the puzzle. We find direct evidence in the EOPP dataset that returns to training increase with job complexity. With heterogeneity in returns, our results cannot be considered structural estimates in the sense of showing the return to training for an average member of the population. Neither can estimated marginal returns be interpreted as the marginal returns to any member of the population. However, under reasonable assumptions our fixed-effect method ensures that the estimates can be interpreted as the average return to a given amount of training for those with that amount of training.

Structural estimation of returns to training when there is heterogeneity presents challenges. 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, Heckman and Vytacil 1998), there are two problems with applying this research to training. First, it is difficult to suggest a plausible instrument. Second, as with measurement error, the mixed continuous-discrete structure complicates the problem. The only paper that we are aware of that deals with a problem of this type is Kenney et al. (1979).[36] We leave more complete analysis of heterogeneity in returns to training as a topic for future research.

[36] Kenney et al. (1979) obtain structural estimates of the return to college education in a model where there is a mass point at zero years of college. In their model, the returns to entering college are heterogeneous, though the returns to years of college conditional on entering are not.

While structural interpretation is difficult due to heterogeneity and measurement error, the size of the returns combined with the short length of most training spells creates a strong presumption that returns must be steeply declining. The returns of 6-8 percent for an initial week of training early in the job match that we find are almost certainly much less in succeeding weeks. One might speculate that the brief but powerful training spells we observe reflect training that is oriented toward teaching how to perform specific job-related tasks rather than broader human capital development. For example, teaching a computer programmer a new programming language may greatly increase his or her productivity in a given job, but training much beyond the basics of the language may have quite small effects. Broader development of programming skills may be left to prior schooling or to job experience. Stepping back from structural interpretation, the logarithmic specification for the returns to formal training stands in stark contrast to specifications compatible with the returns to schooling, which are broadly linear and exhibit increasing returns—“sheepskin effects”--at levels associated with degrees.

References

Abraham, Katharine G., and Henry S. Farber. 1987. "Job Duration, Seniority, and Earnings." *American Economic Review* 77:3, pp. 278-297.

Acemoglu, Daron and Jörn-Steffen Pischke. 1999. "The Structure of Wages and Investment in General Training." *Journal of Political Economy* 107, pp. 539-572.

Altonji, Joseph G. and James R. Spletzer. 1991. "Worker Characteristics, Job Characteristics, and the Receipt of On-the-Job Training." *Industrial and Labor Relations Review* 45:1, pp. 58-79.

Angrist, Joshua, Guido Imbens, and Donald Rubin. 1996. "Identification of Causal Effects using Instrumental Variables." *Journal of the American Statistical Association,* 91, pp. 444-472.

Ashenfelter, Orley and Cecilia Rouse. 1998. "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics,* 113:1, pp. 253-284.

Ashenfelter, Orley and David J. Zimmerman. 1997. "Estimates Of The Returns To Schooling From Sibling Data: Fathers, Sons, And Brothers." *The Review of Economics and Statistics* 79:1, pp. 1-9.

Barron, John M., Mark C. Berger, and Dan A. Black. 1997a. "How Well do We Measure Training?" *Journal of Labor Economics* 15:3 (part 1), pp. 507-528.

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

Barron, John M., Mark C. Berger, and Dan A. Black. 1999. "Do Workers Pay for On-the-Job Training?" *Journal of Human Resources*, pp. 235-252.

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

Bartel, Ann P. 1995. "Training, Wage Growth, and Job Performance: Evidence from a Company Database." *Journal of Labor Economics* 13:3, pp. 401-425.

Becker, Gary. 1975. *Human Capital*, 2nd Edition. Columbia University Press, New York.

Brown, James N. 1989. "Why Do Wages Increase with Tenure? On-the-Job Training and Life-Cycle Wage Growth Observed Within Firms." *American Economic Review* 79:5, pp. 971-991.

Gallant, A. Ronald. 1981. “On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form.” *Journal of Econometrics* 15, pp. 211-245.

Heckman, James. 1997. “Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations.” *Journal of Human Resources* 32:3, pp. 441-462.

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

Heckman, James, and Edward Vytacil. 1998. “Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling when the Return is Correlated with Schooling.” *Journal of Human Resources* 33:4, pp. 974-987.

Kenney, Lawrence W., Lung-Fei Lee, G.S. Maddala, and R. P. Trost. 1979. “Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data.” *International Economic Review* 20:3, pp. 775-789.

Lengermann, Paul A. 1999. “How Long Do the Benefits of Training Last? Evidence of Long Term Effects Across Current and Previous Employers, Education Levels, Test Scores, and Occupations." forthcoming, *Research in Labor Economics*.

Li, Qi, and Thanasis Stengos. 1996. “Semiparametric Estimation of Partially Linear Panel Data Models.” *Journal of Econometrics* 71:1-2, pp. 389-397.

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

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

Loewenstein, Mark A. and James R. Spletzer. 1998. "Dividing the Costs and Returns to General Training." *Journal of Labor Economics* 16:1, pp. 142-171.

Loewenstein, Mark A. and James R. Spletzer. 1999a. "Formal and Informal Training: Evidence from the NLSY." forthcoming, *Research in Labor Economics*.

Loewenstein, Mark A. and James R. Spletzer. 1999b. “General and Specific Training: Evidence and Implications.” forthcoming, *Journal of Human Resources*.

Lynch, Lisa M. 1992. "Private Sector Training and the Earnings of Young Workers." *American Economic Review* 82:1, pp. 299-312.

Mincer, Jacob. 1988. "Job Training, Wage Growth, and Labor Turnover." NBER Working Paper #2690.

Mincer, Jacob. 1989. "Human Capital and the Labor Market: A Review of Current Research." *Educational Researcher* 18, pp. 27-34.

Parent, Daniel. 1999. "Wages and Mobility: The Impact of Employer-Provided Training." *Journal of Labor Economics* 17:2, pp. 298-317.

Pischke, Jörn-Steffen, 1999. "Continuous Training in Germany," mimeo, MIT Department of Economics.

Royalty, Anne Beeson. 1996. "The Effects of Job Turnover on the Training of Men and Women." *Industrial and Labor Relations Review* 49:3, pp. 506-521.

Veum, Jonathan R. 1995. "Sources of Training and their Impact on Wages." *Industrial and Labor Relations Review* 48:4, pp. 812-826.

Table 1
Descriptive Statistics, NLSY

Variable	Mean	Std. Dev.	Min.	Max.
Ln Wage	1.85	0.47	0.00	4.53
# train. spells, current job	0.60	1.33	0.00	21.00
Training Hours, Training > 0	351.59	1070.96	0.50	19200.00
Ln (Training + 1), Training > 0	4.51	1.55	0.41	9.86
Year=1980	0.02	0.14	0.00	1.00
Year=1981	0.03	0.17	0.00	1.00
Year=1982	0.04	0.21	0.00	1.00
Year=1983	0.05	0.22	0.00	1.00
Year=1984	0.06	0.24	0.00	1.00
Year=1985	0.07	0.25	0.00	1.00
Year=1986	0.07	0.26	0.00	1.00
Year=1987	0.08	0.26	0.00	1.00
Year=1988	0.08	0.27	0.00	1.00
Year=1989	0.09	0.28	0.00	1.00
Year=1990	0.09	0.28	0.00	1.00
Year=1991	0.08	0.27	0.00	1.00
Year=1992	0.08	0.27	0.00	1.00
Year=1993	0.08	0.27	0.00	1.00
Year=1994	0.06	0.24	0.00	1.00
Black	0.25	0.43	0.00	1.00
Hispanic	0.18	0.38	0.00	1.00
Age	27.62	4.19	16.00	37.83
Female	0.50	0.50	0.00	1.00
AFQT (residual)	-0.06	20.16	-65.48	45.94
Years education	12.62	2.21	0.00	20.00
Ever married	0.61	0.49	0.00	1.00
Union	0.19	0.39	0.00	1.00
Managerial/prof. (1^{st} yr. in job)	0.17	0.38	0.00	1.00
Other white-collar (1^{st} yr. in job)	0.29	0.46	0.00	1.00
Missing AFQT	0.06	0.23	0.00	1.00
Missing Union	0.06	0.24	0.00	1.00
Any ongoing training	0.03	0.18	0.00	1.00
# spells missing hrs, current job	0.17	0.62	0.00	9.00
n	15,876			
Obs.	61,033			

Table 2
Returns to Training for Different Functional Forms, NLSY

Specification	$\overline{R}^2$	Fraction Fourier Series Explained	Total Effect at Median
Complete Sample			
No Training Vars.	0.1366	--	--
Linear	0.1371	.281	0.001
Quadratic	0.1371	.347	0.002
Square root	0.1378	.729	0.016
Log	0.1381	.906	0.036
# spells	0.1375	.559	0.009
Fourier series	0.1383	--	0.031
n	15,876		
Obs	61,033		
Training Outliers Omitted*			
No Training Vars.	0.1349	--	--
Linear	0.1354	.313	0.002
Quadratic	0.1356	.459	0.006
Square root	0.1361	.721	0.019
Log	0.1364	.896	0.037
# spells	0.1361	.610	0.012
Fourier series	0.1367	--	0.032
n	15,847		
Obs.	60,646		

*Top 1% of training duration, number of spells omitted.

Table 3
Returns to Training in Spline Specification, NLSY

Interval	Slope (x 100)	Difference with previous segment
Complete Sample		
T≤ 40	0.078	
	(0.017)	
40≤T≤78.9	-0.009	-0.086
	(0.024)	(0.037)
78.9≤T≤219.2	0.022	0.030
	(0.007)	(0.029)
219.2≤T	0.001	-0.021
	(0.000)	(0.007)
n	15,876	
Obs	61,033	
$\bar{R}^2$	0.1383	
Training Outliers Omitted[#]		
T≤ 40	0.078	
	(0.017)	
40≤T≤78.9	-0.011	-0.089
	(0.024)	(0.037)
78.9≤T≤219.2	0.026	0.037
	(0.008)	(0.029)
219.2≤T	-0.000	-0.026
	(0.001)	(0.008)
n	15,847	
Obs.	60,646	
$\bar{R}^2$	0.1366	

[#]Top 1% of training duration, number of spells omitted.

Table 4
Lagged Effects of Training on Log Wages, NLSY

Coefficient	1)	2)	3)	4)
Current Log Training	0.0083	0.0057	0.0058	0.0058
	(0.0010)	(0.0011)	(0.0011)	(0.0011)
Log Training Lagged 1 Year		0.0051	0.0036	0.0036
		(0.0009)	(0.0010)	(0.0010)
Log Training Lagged 2 Years			0.0029	0.0031
			(0.0010)	(0.0012)
Log Training Lagged 3 Years				-0.0003
				(0.0012)
Total Effect of Training	0.0083	0.0109	0.0123	0.0122
	(0.0010)	(0.0011)	(0.0012)	(0.0013)

Table 5
Interactions of Training on Current Job with Other Variables, NLSY

Coefficient	1)	2)	3)
Sum* of Log Training Coefficients	0.0136	0.0221	0.0294
		(0.0029)	(0.0042)
	(0.0013)		
Sum of Interactions with Previous Training	-0.0007	-0.0007	-0.0008
		(0.0003)	(0.0004)
	(0.0003)		
Sum of Interactions with Tenure at 0 Tenure		-0.0021	-0.0010
(Quadratic Interaction with Tenure)		(0.0007)	(0.0009)
Sum of Interactions with Exper. at 0 Exper.			-0.0019
(Quadratic Interaction with Experience)			(0.0011)

*Over lags.

Table 6
Increase in Log Wages from 40 Hours of Training for Selected Values of Characteristics, NLSY

Initial Hours of Training, Current Job	Hours of Training on Previous Jobs	Years Tenure	Years Experience	Increase in Log Wages from 40 Hours of Training
0	0	1.19	4.54	0.0801
		2.60	7.44	0.0660
		5.00	10.65	0.0521
	200	1.19	4.54	0.0651
		2.60	7.44	0.0510
		5.00	10.65	0.0372
78.9	0	1.19	4.54	0.0088
		2.60	7.44	0.0072
		5.00	10.65	0.0057
	200	1.19	4.54	0.0071
		2.60	7.44	0.0056
		5.00	10.65	0.0041

Table 7

Descriptive Statistics, EOPP

Variable	Mean	Std. Dev.	Min.	Max.
Ln Wage Growth	0.19	0.21	-1.99	1.95
Formal training indicator	0.13	0.34	0.00	1.00
Informal training indicator	0.95	0.21	0.00	1.00
Hrs. formal tr., formal tr. > 0	75.14	103.76	1.00	640.00
Hrs. informal tr., informal tr. > 0	134.17	177.70	1.00	2070.0
Ln (formal tr. + 1), formal tr. > 0	3.61	1.24	0.69	6.46
Ln (informal tr. + 1), inf. tr. > 0	4.25	1.21	0.69	7.64
Ln # weeks until fully trained	2.23	1.24	0.00	6.033
Years relevant experience	2.33	4.45	0.00	40.00
Rel. experience squared	25.24	110.42	0.00	1600.00
Age	26.57	8.84	16.00	64.00
Years education	12.47	1.66	2.00	24.00
Union	0.11	0.31	0.00	1.00
Ln establishment size	2.89	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
Other white-collar	0.56	0.50	0.00	1.00
Obs.	1,550			

Table 8

Returns to Aggregate Training for Different Functional Forms, EOPP

Specification	$\bar{R}^2$	Total Effect at Median
Complete Sample		
No Training Vars.	0.0714	--
Linear	0.0833	0.009
Quadratic	0.0838	0.014
Square root	0.0877	0.039
Log	0.0864	0.081
Obs.	1,550	
Training Outliers Omitted*		
No Training Vars.	0.0694	--
Linear	0.0753	0.009
Quadratic	0.0830	0.031
Square root	0.0803	0.035
Log	0.0813	0.074
Obs.	1,533	

*Top 1% of training duration omitted.

Table 9

Returns to Formal and Informal Training for Various Log Specifications, EOPP

Coefficient	Unweighted Sum	Cobb-Douglas	Cobb-Douglas Plus Formal-Informal Interaction	CES	CES Plus Aggregate Training - Number of Weeks until Qualified Interaction
Log Unweighted Aggregate Training	0.019 (0.004)				
Log Formal Training		0.016 (0.004)	0.021 (0.017)		
Log Informal Training		0.015 (0.004)	0.015 (0.004)		
Log Formal Training x Log Informal Training			-0.001 (0.003)		
Log Weighted Aggregate Training				0.029 (0.005)	0.010 (0.009)
Log Weighted Aggregate Training x Log Number Weeks Until Fully Qualified					0.007 (0.003)
Weight on Formal Training				0.950 (0.028)	0.938 (0.035)
Obs.	1,550	1,550	1,550	1,550	1,550
$\bar{R}^2$	0.0864	0.0922	0.0917	0.0952	0.0969

Table 10

Returns to Formal and Informal Training for Different Functional Forms, EOPP

Specification	$\overline{R}^2$	Fraction Fourier Series Explained	Weight on Formal Training	Total Effect of Formal Training at Median Positive Value	Total Effect of Informal Training at Median Positive Value
Complete Sample					
Linear	0.0861	0.597	0.80	0.015	0.007
Quadratic	0.0882	0.699	0.87	0.031	0.009
Square root	0.0930	0.905	0.89	0.039	0.031
Log	0.0952	0.896	0.95	0.062	0.038
Fourier Series	0.0945	--	0.90	0.052	0.052
Obs.	1,550				
Training Outliers Omitted*					
Linear	0.0785	0.0576	0.84	0.014	0.006
Quadratic	0.0824	0.0710	0.89	0.039	0.014
Square root	0.0852	0.0851	0.91	0.034	0.028
Log	0.0883	0.0861	0.95	0.051	0.037
Fourier Series	0.0890	--	0.91	0.055	0.050
Obs.	1,533				

*Top 1% of (unweighted) training duration omitted.

Figure 1. Estimated Effect of Hours of Training on Ln Wages for Different Specifications--NLSY

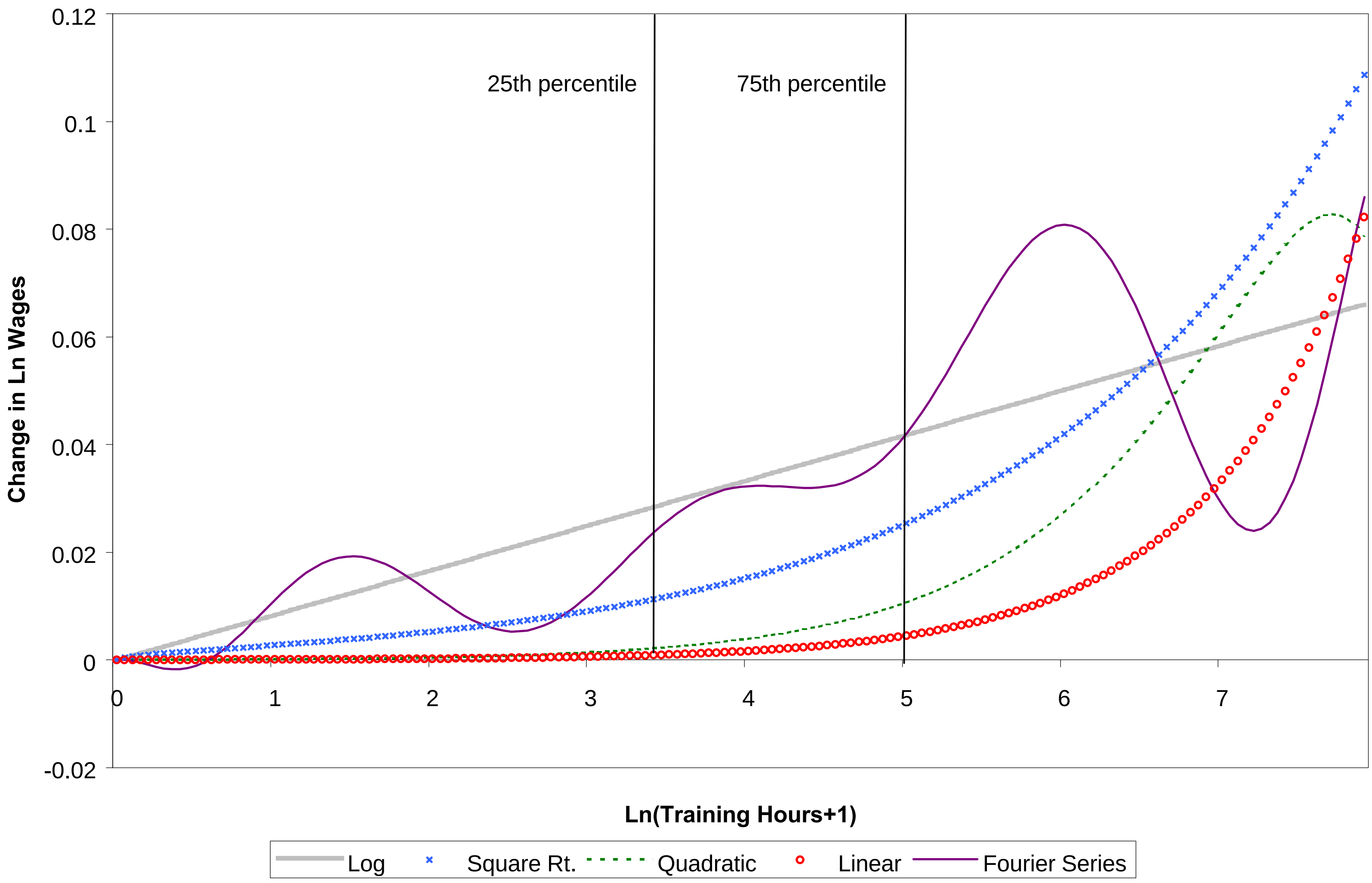

Figure 2. Estimated Effect of Formal Hours of Training on Ln Wages for Different Specifications--EOPP

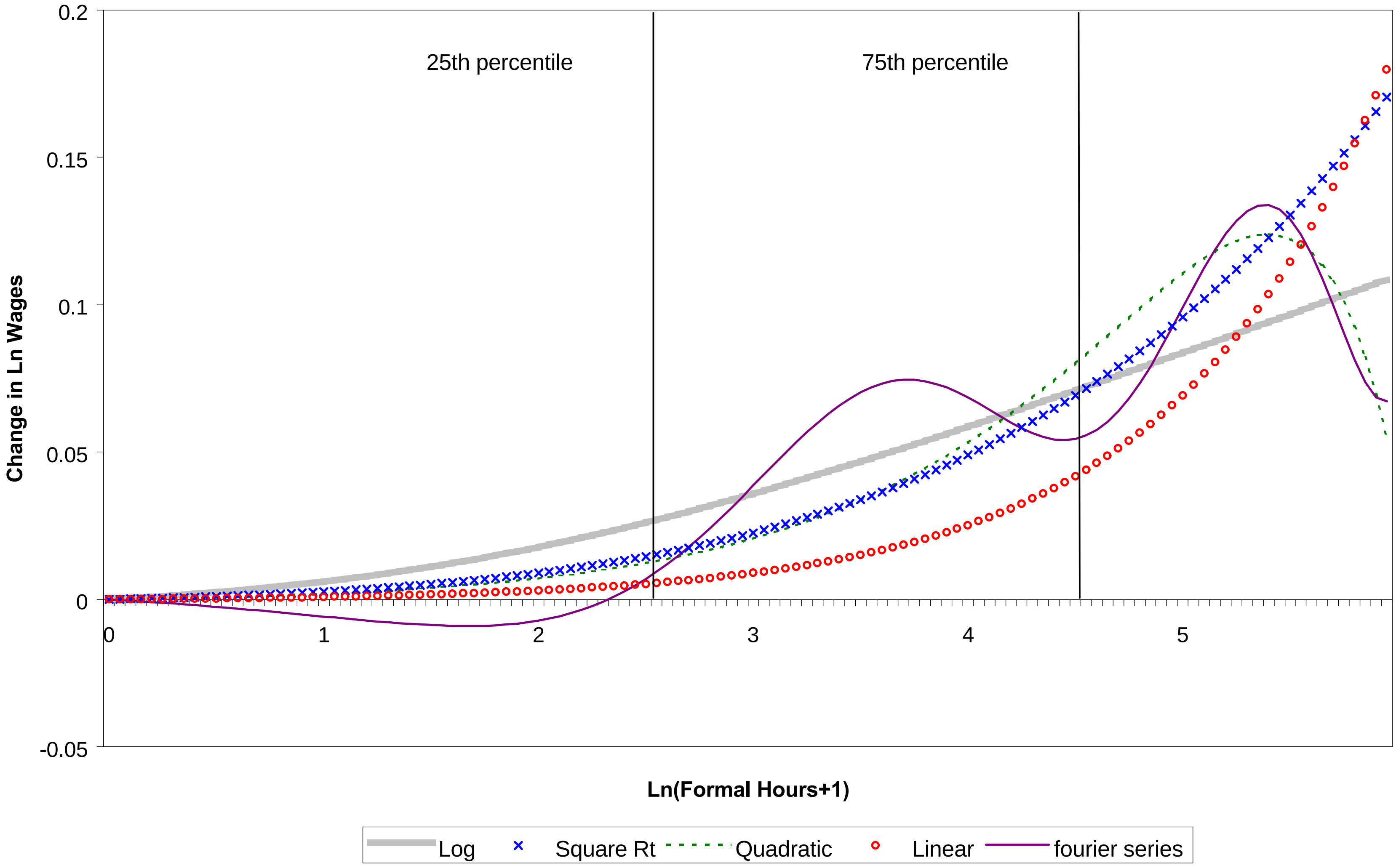

Figure 3. Estimated Effect of Informal Hours of Training on Ln Wages for Different Specifications--EOPP

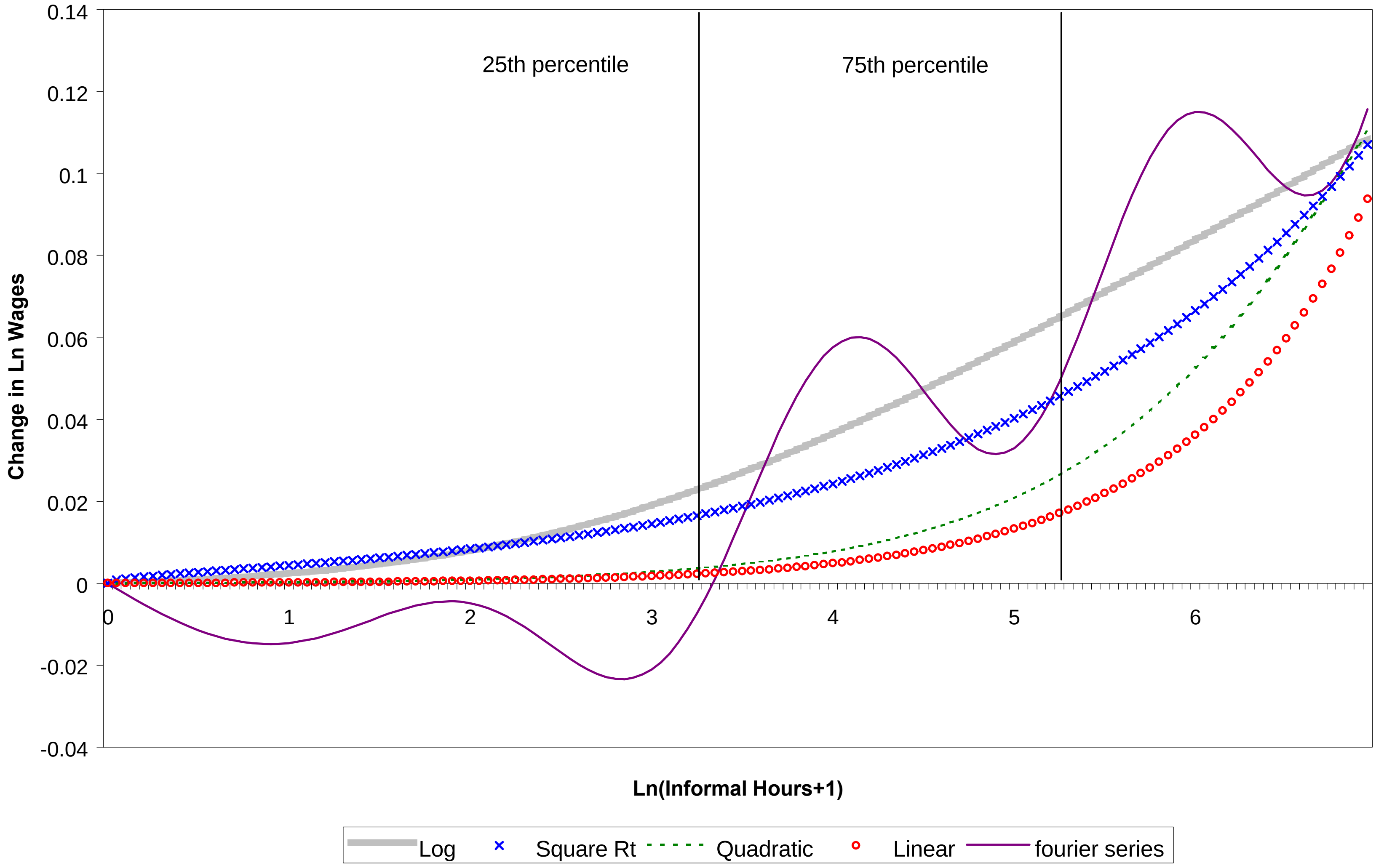

Figure 4