BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF EMPLOYMENT AND UNEMPLOYMENT STATISTICS

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

Harley Frazis, U.S. Bureau of Labor Statistics
Mark A. Loewenstein, U.S. Bureau of Labor Statistics

Working Paper 367
July 2003

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: 202-691-7394
Frazis_H@bls.gov

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

Phone: 202-691-7385
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 and at seminars at Virginia Commonwealth University and George Washington University 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 estimates the wage returns to training, paying careful attention to the choice of functional form. Both the National Longitudinal Survey of Youth (NLSY) and Employer Opportunity Pilot Project (EOPP) datasets indicate that the return to an extra hour of formal training diminishes sharply with the amount of training received. A cube root specification fits the data best, but the log specification also does well. The linear and quadratic specifications substantially understate the effect of training.

If wages are not adjusted continuously, estimating the total effect of training requires that one include lagged and lead training as well as current training in the regression equation. Consequently, the NLSY is ideally suited to estimate the total return to training. We find very large returns to formal training. These returns are sharply reduced when one adjusts for heterogeneity in wage growth. Returns are reduced further when one takes into account the effect of promotions and the fact that direct costs are a substantial portion of the total cost of training. 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 relatively small.

After correcting for confounding factors, we are left with a return to training that is several times the returns to schooling. Heterogeneity in returns explains how returns to formal training can be so high while most workers do not get formal 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."

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 makes comparisons difficult. 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

---

[1] A non-exhaustive list of references here includes Altonji and Spletzer (1991), Barron, Berger, and Black (1999), Barron, Black and Loewenstein (1989), Bartel (1995), Brown (1989), Lengermann (1999), Lillard and Tan (1986), Loewenstein and Spletzer (1996, 1998, 1999a), Lynch (1992), Mincer (1988), Pischke (2001) and Veum (1995).

[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 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

related Small Business Administration (SBA) data have generally used log specifications (for example, see Barron, Black, and Loewenstein 1989 and Barron, Berger, and Black 1997b), 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).

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--over 150 percent for typical values of training-- and incidence low (less than 25% of job spells in our NLSY sample have any formal training) we consider possible explanations for the seeming contradiction between high rates of return and low incidence of training.

Most of our analysis relies on NLSY data. Of the various datasets with training information, the NLSY provides the best information on formal training at all levels of tenure. In our discussion of how to estimate rates of return to training, we demonstrate that under reasonable assumptions, wage regressions must include lagged and lead training as well as current training in order to produce a rate of return estimate. The NLSY, which is longitudinal, is thus ideally suited to estimate the return to training. We supplement our NLSY analysis with estimates based on EOPP data. EOPP is an

depreciation and 26 percent with 20 percent deprecation. Interestingly, Lengermann (1999) finds no evidence that the return to training depreciates with time.

employer survey and provides the best information on formal and informal training at the start of the job. The EOPP data enable us to check that our key NLSY results are not distorted by our inability to control for informal training or by the specific survey methods used in the NLSY.

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. We find that a cube root specification fits the data best, but the log specification also does well. These specifications, along with our semi-nonparametric estimator, indicate that there are very substantial returns to the initial interval of formal training. In contrast, in both datasets the linear specification substantially understates the effect of training, and in the NLSY the quadratic also substantially understates the effect of training.

In section III, we develop a simple theoretical framework that allows us to estimate the rate of return to training. Our estimate of the rate of return is very high. Rates of return at median positive training of 60 hours are in the 150-175 percent range for the better-fitting parsimonious specifications and over 180 percent for the Fourier series. These numbers are much higher than, for example, estimates of returns to schooling, and present a puzzle in view of the fact that only a minority of respondents in either the NLSY or EOPP have any formal training.

In section IV, we discuss five potential explanations as to why estimated returns to training are so high: heterogeneity in wage growth, measurement error, promotions, direct costs of training, and heterogeneity in returns to training. Using the cube root specification as our base, we show how correcting for these various factors causes one to adjust the estimated return to training.

The estimated return to training is sharply reduced when one adjusts for heterogeneity in wage growth, a result similar to that obtained by Pischke (2001). Returns are reduced further when one takes into account the effect of promotions and the fact that direct costs are a substantial portion of the total cost of training. 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 relatively small.

After correcting for confounding factors, we are still left with a return to training that is several times the returns to schooling. Heterogeneity in returns explains how returns to formal training can be so high while most workers do not get formal training. Both the EOPP data and the NLSY provide evidence of such heterogeneity: in EOPP, the return to training is significantly higher in more complex jobs, and in the NLSY managers and professionals have higher returns to training than do other occupations. If, as seems reasonable, part of the heterogeneity in returns is due to unobservable factors, our estimates can be regarded as the return to training for the trained, but cannot be extrapolated to the untrained.

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 training and log wages. In view of the advances over the last several decades in the theory and computational feasibility of non-parametric and semi-parametric estimators, it might be questioned whether this is a useful task. Why not characterize the relationship between training and wages nonparametrically? We think that it is useful to recommend a particular parsimonious specification for applications where one might not have the luxury of estimating a more flexible functional form. For example, one might have a specification where training is interacted with other variables or contains lead and lag terms such that estimating a more flexible function form is not practical. (We present an example of this in Section 4.) Alternatively, training may not be the focus of the analysis, but one may want to control for the effect of training on wages when estimating other effects. For example, Altonji and Pierret (2001) seek to correct for the effect of training when examining the effect of cognitive skills on wage growth.

NLSY Data

The NLSY is a dataset 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 over 90 percent for each year until 1996, and as of 2000 were 83 percent. We use data from the 1979 through 2000 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

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

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]

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]

The focus of our analysis is training whose explicit cost is at least partly paid for by the employer.[7] 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

---

[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.

[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] Naturally, this cost can be passed on to the worker in the form of a lower wage. Eighty-fiver 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. We use this framework when discussing rates of return to training in the next section.

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 non-company 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.[8] 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. This variable is obtained by adding the training a worker has completed in the current year to the training he has received in all previous years on the current job.

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

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

ε_{ijt} is a mean zero error term, uncorrelated with X_{ijt}. All specifications are run as 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,[9] 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), non-employer-paid training,[10] and dummies for ever married, part-time, enrolled in school, missing part-time, and year dummies are also included. As additional controls 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.

---

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

[10] A preliminary functional form analysis showed that a linear specification was best for non-employer-paid training.

Descriptive statistics are shown in Table 1. Table 2 gives more detail on the distribution of training by showing selected percentiles of the positive distribution of both the stock of training and of training during the previous year. Note that the distribution is quite skewed to the right. Log-normality appears to be a good approximation of the distribution. Using the Box-Cox transformation $B(T; \lambda) \equiv (T^{\lambda} - 1)/\lambda$ as a transformation of the distribution of T to normality (where λ=0 corresponds to a log-normal distribution and λ=1 to a normal distribution; see Greene 2000, p. 444-46), the estimated value of λ for the positive training sample is .03 for the stock of training and -.01 for training in the previous year.

As an aid to determining the best simple functional form, we make a different use of the Box-Cox transformation by first estimating a model where $f(T) = B(T;\lambda)$. Values of λ of 1 and 0 correspond to $f(T) = T$ and $f(T) = \ln(T)$, respectively. The estimation was done by non-linear least squares. The estimated value of λ is .350 (bootstrap standard error .062), very close to the value of 1/3 corresponding to the cube root.

The results for different functional forms for training are shown in Table 3.[11] 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 $\bar{R}^2$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

[11] Estimating regressions for men and women separately gives results quite similar to Table 3.

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 $\overline{R}^2$ over 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 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 non-linear, 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 non-linearity.[12]

To contrast the effects of training on wages implied by the different functional forms, the last column in Table 3 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, over 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.

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

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

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 semi-nonparametric 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).[13]

In practice, linear and quadratic terms are usually added to the expansion. Moreover, for non-periodic 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

[13] 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.

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 2**π**. 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 *K*th 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 1 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 was 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 series and f(T) relative to a specification which omits training. As can be seen in the third column of Table 3, the cube root specification is closest to the Fourier series, and the linear specification is the furthest. Indeed, the cube root specification explains over 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.

---

[14] However, asymptotic optimality is not proven if observations are not independent. We are unaware of equivalent results for this case.

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 volatile nature 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.

Why do the linear and quadratic functional forms track the Fourier series so poorly, especially 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. (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 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.

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

EOPP

The NLSY provides strong evidence that returns to formal training decline greatly with the quantity of such training. As a check on these 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] We will make use of this measure later on in our analysis.

EOPP's 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.

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

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

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. 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 had received any vocational training or belonged to a union. In addition, we include the logarithm of the number of employees at 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 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 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 4. 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,

---

[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

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 .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 (.291)).

The results for the various functional forms for training are shown in Table 5. 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 too much weight on the Fourier results. Indeed, the simple cube root and quadratic specifications both have a higher $\overline{R}^2$ than the Fourier series. The volatility of the results of the quadratic specification between samples makes us reluctant to recommend it as an alternative to the cube root. In addition to the

---

job. Since wages increase over time, tenure is positively correlated with wage growth.

sensitivity to outliers in the NLSY, the predicted effect of training at the EOPP median positive training of 38 hours using a quadratic specification is three times higher in EOPP than in the NLSY if one omits outliers (and nine times higher using the complete sample). For the log and cube root specifications the effects estimated using EOPP are only about 50 percent higher than those from the NLSY.[18]

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 U.S. (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.

<u>Rate of Return Calculations Using Coefficients on Lagged, Current, and Lead Training</u>

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

[18] As shown below, one year may not be sufficient to capture the full effect of training in the NLSY. When one allows one lead and one lag, the estimated return to training in the NLSY is substantially closer to the EOPP estimated return, which is itself estimated over a two year period . Another possible explanation for the higher estimated return in EOPP is that workers receiving greater amounts of training during the first three months may also receive more training during the rest of the two year period. Consistent with this, Loewenstein and Spletzer (1996) find that within the NLSY, training incidence is highly correlated within jobs.

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 $MR(T) \equiv \frac{d\ln(g(T))}{dT} \equiv \frac{g'(T)}{g(T)}$. 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 sub-optimal investment in training. In contrast, assuming that our fixed-effect estimates of returns can be interpreted structurally, high marginal returns would imply sub-optimal investment in training. We return to this issue below.

Note also that high observed average IRR's 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 realize all the gain to training -- because, say, training is general (see Becker 1975). 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 and ends between t=1 and t =2. That is, letting τ_1 be the date that training starts and τ_2 be the date that training ends, $1 < \tau_1 < \tau_2 < 2$. Then the worker's wage is $g(0)$ during the first year of employment ($0 \leq t < 1$) and $g(T)$ during the third year ($2 \leq t < 3$). 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 remaining time 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. Taking first differences one obtains $\ln(w_2) - \ln(w_1) =$

$\ln(\pi_1 g(0) + \pi_2 g(T)) - \ln(g(0))$ and $\ln(w_3) - \ln(w_2) = \ln(g(T)) - \ln(\pi_1 g(0) + \pi_2 g(T))$.

Note that the effect of training will be spread over two periods - the period of training itself and the period after training has been completed. Note also that if π_2 is sufficiently small relative to the time spent in training, $\ln(w_2) - \ln(w_1)$ will be negative.

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. How should training enter the regression? To answer this question, 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, so that $\ln(w_{t2}) = \ln(\pi_1 g(0) + \pi_2 g(T_2))$ and $\ln(w_{t3}) = \ln(g(T_2))$. That is, the training T enters the wage equation at time t_2 as *current* training and at time t_3 as *lagged* training. On the other hand, if training is completed after the survey date t_2, then $T_2 = T_1 = 0$ and $T_3 = T$, so that $\ln(w_{t2}) = \ln(\pi_1 g(0) + \pi_2 g(T_3))$ and $\ln(w_{t3}) = \ln(g(T_3))$. In this case, the training T enters the wage equation at time t_2 as *lead* training and at time t_3 as *current* training.

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.[19]

The foregoing has assumed that the worker bears all the costs (in terms of foregone production) and obtains all the returns to training. If the training is to some extent firm-specific, or if there are frictions in the labor market that cause the firm to share in the cost of general training [Loewenstein and Spletzer 1998, Acemoglu and

[19] 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.

Pischke 2001], then the wage effect will underestimate the return to training in terms of productivity.[20] The observed wage effect is thus a lower bound (subject to caveats explained later).

<u>Rate of Return Results.</u>

Table 6 shows results for the specifications considered in Table 3, 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 3, 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 $IRR(T) = \frac{2000\sum_{-1}^{1}\beta_t f(T)}{T}$, where β_{-1} is lagged, β_0 is current, and β_1 is lead training. Rates of return at median positive training of 60 hours are in the 150-175 percent range for the better-fitting parsimonious specifications and over 180 percent for the Fourier series. 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 mean returns 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. (For the parametric estimators, the

---

[20] Other contracting situations are plausible. For example, the costs of training may be shared by workers

estimated mid-range returns are similar to the estimates at the median.) The estimated mid-range returns, shown in the first two rows of Table 7, are in the neighborhood of 180 percent, comparable to that for the log specification.

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 hours of training. The effect of training on wages in EOPP is of a comparable order of magnitude. These numbers are much higher than, for example, estimates of returns to schooling. The numbers also present a puzzle in view of the fact that only a minority of jobs in both the NLSY and EOPP have any formal training--13% in EOPP in the first three months, and about 25 percent in the NLSY as of the last observation on the job. Taking the results literally, it would appear that potentially profitable investments in training are not being made. In this section we discuss five potential explanations as to why estimated returns to training are so high: heterogeneity in wage growth, measurement error, promotions, direct costs of training, and heterogeneity in returns to training.

Heterogeneity in Wage Growth

Our fixed-effect regressions control for all factors whose effect on wages remains unchanged during a job match. However, 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 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

---

who do not receive training. We leave these considerations for another paper.

NLSY wage equation.[21] (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 coefficients on lead, current and lagged training should fall.

This is in fact what we observe. The third and fourth rows of Table 7 shows mid-range rates of return for the Fourier series.[22] 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. The interaction coefficients imply that respondents who end up with 60 hours of training average about .8 percent per year more rapid wage growth initially and about .6 per year after 2.5 years.

When the (cube root of) final training interactions are added, the differences in fit among the various functional forms become smaller, as shown in Table 8. The Q^2 statistic favors the log and cube root specifications. The rate of return for the Fourier series is higher than that for the parametric specification both at the median itself and between the 25th and 75th percentiles. Table 9 shows the coefficients for the leads and lags in the cube root specification, with and without final training interactions. Note that the lag and lead coefficients decline greatly in magnitude.

The identification of the final-training/tenure interaction coefficients merits closer examination. These coefficients are identified because interactions of stocks of (current,

---

[21] Pischke (2001) controls for unobserved wage growth heterogeneity by including an interaction between tenure and an individual fixed effect in a wage regression that already includes a (noninteracted) fixed-effect term. Our approach is more flexible in that it allows for a more 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. Pischke finds that adding the fixed-effect tenure interaction sharply reduces the coefficient on a training dummy, but not on training duration, suggesting that high growth individuals select into short duration training spells.

[22] As for the specification without final training interactions, K equals 13 for the complete sample and 2 for the outlier-omitted sample.

lead, and lagged) training with tenure are excluded from our regression, as are leads and lags beyond one period. In particular, training led *M-t* periods, where *M* is the year an individual leaves his current job, is precisely final training, so that interactions of training led *M-t* periods with tenure will be equal to final training times tenure.

Excluding long leads of the stock of training is in our judgment clearly appropriate. This still raises the question of whether interactions of the stock of training with tenure should enter the wage equation. One might expect that job tenure and training would to some extent be substitutes, so that training that occurs after long tenure with the employer would have lower returns. We use the cube root specification to examine this question, as Fourier series estimates are quite imprecise. Results are shown in Table 10 for the outlier-omitted sample (results from the full sample are similar). For comparison purposes, column 1 reports the results of the cube root specification with the final training interactions (i.e., the specification reported in column 4 of Table 9).

Column 2 shows what happens when one adds interactions of tenure with current, lead, and lagged training. The point estimates indicate that returns to training decline with tenure, especially in the first few years, as one would expect if job training were a substitute for on-the-job learning. However, the interactions are imprecisely estimated; the joint p value of the interactions is 0.57. In the bottom half of Table 10 we show the derivative of the effect of training (at the median of 60 hours) with respect to tenure at various points in the distribution of tenure. None of the derivatives examined are statistically significant at conventional levels.

Attempting to simultaneously estimate leads and lags for training, final training/tenure interactions, and interactions of lead and lagged training with tenure may well be demanding more from our data than they can reasonably be expected to show. The small size of the lead and lagged coefficients in Table 9 is another indication that it may be overly ambitious to attempt to estimate interactions of lead and lagged training with tenure. Accordingly, Column 3 of Table 10 shows the results of interacting current training with tenure, but not lead or lagged training. Here there is stronger, though not overwhelming, evidence that the effect of training declines with tenure. The *p* value of the interactions declines to 0.12, and the slope of the effect of training with respect to tenure at the 25th percentile of tenure is substantial and significant at the 5 percent level. The results suggest a decline in the returns to training with increased tenure in the first few years on the job.

While the estimated return for the median tenure is similar whether or not tenure interactions are included, the inclusion of interactions with tenure increases the estimated return for low values of tenure. Using the cube root specification, the return at the 25th percentile of tenure and 60 hours of training is 123 percent. This compares to a 95 percent return when the tenure interaction is omitted . Both of these estimates are somewhat conservative given that the Fourier series returns are higher.[23]

In summary, heterogeneity in wage growth is responsible for a significant part of the apparent high returns to training.[24] Furthermore, it is difficult to estimate both final training effects and tenure interactions with the stock of training, but there is evidence that the returns to training decline with tenure. After correcting for wage growth

[23] The inclusion of interactions of tenure with current training also increases the effect of the final-training interactions: aside from the increase in wages immediately after being trained, respondents who end up with 60 hours of training are estimated to have 1.6 percent higher wage growth at the beginning of the job and 0.8 percent higher wage growth after 2.5 years.

heterogeneity, estimated returns to training for workers with the median positive value of training are in the neighborhood of 125 percent; returns for workers with low tenure are likely higher.

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. As explained below, this mixed continuous-discrete structure implies that estimates of the effect of short spells of training may 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 $g'' \leq 0$. Since we do not observe true training, the data do not reveal the function g, but instead reveal 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.)

One can distinguish between two types of measurement error: misclassification of training and error in the duration of spells that are classified correctly. Misclassification

[24] A previous draft of this paper found no effect of the final training-tenure interactions on returns to

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 2. 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.

As positive training is approximately log-normal, the simplest assumption is that duration error is log-normal. We show in the Appendix, that if the return to training is a linear function of $\ln(T^*)$, then the effects of duration error will cancel out at the geometric mean. If the return to training declines at a slower (faster) than logarithmic rate, then duration error will cause the return to training to be biased upward (downward). Using parameters from the NLSY and assuming a 2 period structure to the data, we show in the Appendix that under reasonable assumptions, the maximum proportional bias in the

---

training. The difference in results is due to the addition of tenure cubed to the specification, and improved measurement of final training.

estimated return to geometric mean training from duration error is less than 20 percent. Monte Carlo evidence establishes that any positive bias is likely to be smaller in longer panels. Taking into account plausible magnitudes of classification error (also discussed in the Appendix), downward bias from classification error should offset any upward bias from duration error.

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 been increased since the last interview. Respondents also were asked whether they had 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 1987-1990 and 1994-2000 and estimate a wage equation over both sub-periods (where the job is unchanged, separate fixed-effects are estimated in each sub-period).[25] We find that adding the change in responsibilities 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. As reported in Table 7, one is thus left with a rate of return of 75 percent. Similarly, in the cube root specification, adding the change in responsibilities variable reduces the sum of the training coefficients by .0030, implying a reduction in the estimated rate of return of about 39 percentage points. Applying this to the results in Table 8 produces 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 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 clearly have an identification problem; while giving an able worker more responsibilities may increase productivity in the absence of training, a worker's improved ability to carry out more advanced job duties should properly be considered to be part of the return to the training investment. 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 estimated effect of training is clearly too large.

[25] 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

A more reasonable way of accounting for promotions when estimating the return to training is to control for promotions when estimating the effect of the current stock of training, but not when estimating the effect of lagged training-that is, we 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 in effect 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. We have added an indicator variable that takes on a value of 1 when a worker has received a promotion within two years of being hired as an additional explanatory variable. While the coefficient on the promotion dummy is positive and significant, there is virtually no effect on the formal training coefficient.

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, wages and salaries of trainers, payments to outside trainers, tuition reimbursements, and contributions to training funds

---

responsibilities question or b) indicates that his responsibilities have changed as a result of a promotion.

totaled $300 per employee in 1994. The survey also estimated that wages and salaries paid to employees while in formal training totaled $224 over the period May-October 1995 (Frazis et al. 1997). Pro-rating the wage and salary 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.

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. Since the skills required for different jobs are heterogeneous, it makes sense that the returns to training differ across jobs. Both the NLSY and the EOPP data provide 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 11. 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 required effects of training on productivity to make investing in a part-time employee worthwhile.[26]

EOPP contains a variable that may more directly reflect training requirements for a job. Recall that one of the control variables in our wage growth regression is 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, which we refer to as "job complexity". Consistent with this interpretation, "job complexity" is positively related to wage growth, as can be seen in column 1 of Table 12, which reports the key coefficients on training and job complexity in our preferred EOPP specification (i.e., the cube root specification in Table 5).

Column 2 shows the effect of interacting formal training and job complexity and column 3 shows the effect of also including an informal training interaction. The noninteracted training coefficients fall when the job complexity-training interactions are included, and the job complexity – training interactions are both positive. But the standard errors are very high. Apparently, attempting to estimate the separate effects on wage growth of formal training, informal training, job complexity, and job complexity-training interactions places too great demands on the EOPP data. We therefore look for a more parsimonious specification.

We begin by considering a specification of the form:

$$E(\Delta w_i | X, T_{formal}, T_{informal}) = \ln(\psi(T_{formal}, T_{informal})) + X_i\beta \,,$$

[26] In spite of the huge effect of part-time training on wages (60 hours of part-time training raises the wage by 6.7 percent), the rate of return to training part-time blue collar workers is slightly less than the return to training full-time managers and professionals if one calculates the rate of return using 1000 hours for a work year instead of 2000 hours.

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

The perfect substitute specification corresponds to $\rho = -1$, and the Cobb-Douglas specification corresponds to the limit of ψ as $\rho \rightarrow 0$. The perfect substitute specification turns out to be best fitting, although the parameter ρ is very imprecisely estimated. Having determined that the perfect-substitute specification is reasonable, we next estimate a Box-Cox model of the form

$$E(\Delta w_i\,|X, T_{formal}, T_{informal}) =(1/\lambda)(b(T_{formal}+1) +(1-b)(T_{formal} + 1))^{\lambda}-1 + X_i\beta$$

The estimated value of λ is .02, indicating that the log is a good choice for functional form when one aggregates formal and informal training. The results for the log – perfect substitute specification are reported in column 4 of Table 12.[27] Finally, column 5 shows the results of interacting aggregate training with job complexity. The coefficient on the interacted variable is positive and quite large. The return to aggregate training increases by 50 percent going from the 25th to the 75th percentile of weeks until fully qualified. In the presence of measurement error in both variables, this is likely to be a severe understatement of the effect of job complexity on returns to training.[28]

The EOPP and NLSY results provide evidence that the return to training varies greatly across jobs. 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

---

[27] Note that in obtaining aggregate training, the estimated weight on formal training is .935 and the weight on informal training is only .055.

[28] Barron, Berger, and Black (1997a) find substantial discrepancies between employer and employee reports of weeks until fully qualified. Griliches and Ringstadt (1970) demonstrate that measurement error is likely to more severely bias downward the magnitude of the coefficient of a quadratic term than a linear term where the true model is quadratic. For similar reasons, measurement error in both hours of training

wage levels by means of the fixed effect. To see this, consider the following simplified wage model that abstracts from covariates other than training:

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

where $E(e_{it}) = E(\alpha_i) = 0$, $E(\beta_i) = \overline{\beta}$, 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,[29] presumably because their cost of training is lower and/or their return to training is higher. If the cost of training is lower for more able individuals in more productive jobs, that is, if $\text{cov}(\alpha, T) > 0$, 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. Using this data, Loewenstein and Spletzer (1999b) demonstrate that hours of aggregate training are strongly positively correlated with job complexity. And, as noted above, the return to training is higher for individuals who are in more complex jobs.

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 (which, in this case, is equivalent to first differences) is given by:

---

and number of weeks until fully qualified severely downward biases the estimated magnitude of their interaction.

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

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

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

$$= \mathrm{E}(\beta_i|\, T_{i2}= T_0)\varphi(T_0).$$

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 $\bar{\beta}\,\varphi\,(T_0)$, but, as is clear from (10), consistently estimate the effect of a given amount of training *for those with that amount of training*.[30] 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 $\mathrm{E}(\beta_i|T=T_0) >$ $\mathrm{E}(\beta_i|T=0)$: 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

$$(11)\qquad f'(T_0) = \mathrm{E}(\beta_i|T=T_0)\varphi'(T_0) + \frac{\partial \mathrm{E}(\beta_i\,|\,T=T_0)}{\partial T}\varphi(T_0),$$

[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. As with measurement error, 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.

and which will exceed $E(\beta_i|T=T_0)\varphi'(T_0)$ if $\frac{\partial E(\beta_i \mid T = T_0)}{\partial T} > 0$: 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.

Summary

To summarize our discussion: 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 returns in the neighborhood of 40-50 percent at the median positive level of training.[31] Measurement error likely leads to either minimal overestimates or to underestimates at this level of training. These returns are several times the returns to schooling and are very likely an underestimate in that they do not reflect cost-sharing with the employer. Returns appear to be higher for those with low tenure. Heterogeneity in returns is a potential explanation as to how returns to formal training can be so high while most workers do not get formal training. While 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

[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

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 quite volatile.

Our best fitting specifications indicate that there are very substantial returns to the initial interval of formal training. One explanation is heterogeneity in wage growth. Fixed-effect regressions control for all factors whose effect on wages remains unchanged during a job match. However, unobserved factors that affect both wage *growth* and training will bias fixed-effect estimates of the return to training. Controlling for heterogeneity in wage growth by adding interactions of tenure and an individual's final observed training in the current job to the wage equation has the effect of sharply reducing the estimated return to training.

Returns are reduced further when one takes into account the effect of promotions and the fact that direct costs are a substantial portion of the total cost of training. After correcting for confounding factors, we are still left with a rate of return in the neighborhood of 40-50 percent at the median positive level of training. This estimated return, which is several times that associated with schooling, is an underestimate since it does not take into account cost-sharing with the employer.

Heterogeneity in returns explains how returns to formal training can be so high while most workers do not get formal training. Both NLSY and EOPP data provide evidence of such heterogeneity: the return to training is significantly higher for managers and professionals in the NLSY and in more complex jobs in EOPP. With heterogeneity in returns, our results cannot be considered structural estimates in the sense of showing

---

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.

the return to training for an average member of the population, nor 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 estimated average return can be interpreted as the 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 find 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).[32] We leave a more complete analysis of heterogeneity in returns to training as a topic for future research.

[32] 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.

References

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 Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116, pp. 313-350.

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.

Andrews, Donald W.K. 1991. "Asymptotic Optimality of Generalized CL, Cross-validation, and Generalized Cross-validation in Regression with Heteroskedastic Errors." *Journal of Econometrics* 47, 359-377.

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.

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* 34:2, 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.

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, edited by Gary Libecap, (Greenwich, Ct.: JAI Press, 1997), pp. 47-82.

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.

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

Griliches, Zvi, and Vidar Ringstadt. 1970. "Error-in-the-Variable Bias in Nonlinear Contexts." *Econometrica* 38:2, pp. 368-370.

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 Solomon Polachek. 1974. "Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship." *Journal of the American Statistical Association* 69: 346, pp. 350-354.

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.

Jaeger, David A. 2003. "Estimating the Returns to Education Using the Newest Current Population Survey Education Questions." *Economics Letters* 78:3, pp. 385-394.

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." *Research in Labor Economics* 18, pp. 439-461.

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." *Research in Labor Economics*. vol. 18, pp. 403-438.

Loewenstein, Mark A. and James R. Spletzer. 1999b. "General and Specific Training: Evidence and Implications.", *Journal of Human Resources*, 34:7, pp. 710-733.

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. 2001. "Continuous Training in German." *Journal of Population Economics* 14, pp. 523-548.

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	1.88	0.49	0.00	4.53
# train. spells, current job	0.54	1.34	0.00	21.00
Training Hours	40.60	209.14	0.00	9260.00
Ln (Training + 1)	1.02	1.90	0.00	9.13
Training Hours, Training > 0	164.60	396.11	0.50	9260.00
Ln (Training + 1), Training > 0	4.12	1.39	0.41	9.13
Years Tenure	3.80	3.68	0.00	22.77
Year=1980	0.02	0.13	0.00	1.00
Year=1981	0.03	0.16	0.00	1.00
Year=1982	0.04	0.19	0.00	1.00
Year=1983	0.04	0.20	0.00	1.00
Year=1984	0.05	0.22	0.00	1.00
Year=1985	0.06	0.23	0.00	1.00
Year=1986	0.06	0.24	0.00	1.00
Year=1987	0.06	0.24	0.00	1.00
Year=1988	0.07	0.25	0.00	1.00
Year=1989	0.07	0.26	0.00	1.00
Year=1990	0.07	0.26	0.00	1.00
Year=1991	0.07	0.25	0.00	1.00
Year=1992	0.07	0.25	0.00	1.00
Year=1993	0.07	0.25	0.00	1.00
Year=1994	0.06	0.24	0.00	1.00
Year=1996	0.06	0.24	0.00	1.00
Year=1998	0.06	0.23	0.00	1.00
Year=2000	0.05	0.21	0.00	1.00
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/prof. (1st yr. in job)	0.17	0.38	0.00	1.00
Other white-collar (1st yr. in job)	0.34	0.47	0.00	1.00
Part-time	0.12	0.32	0.00	1.00

[Continued]

Table 1, continued

Variable	Mean	Std. Dev.	Min.	Max.
Hours outside training on current job	18.47	165.59	0.00	5440.00
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
# spells missing tr. hrs, current job	0.01	0.12	0.00	4.00
Ever training spell, current employer	0.25	0.43	0.00	1.00
Any training spell, current period	0.10	0.30	0.00	1.00
N	17,809			
Obs.	75,698			

Table 2

Percentiles of Distribution of Hours of Training, Training > 0

Percentile	Cumulative Stock of Training	Hours of Training During Previous Year
10	8	4
25	26	10
50	60	40
75	146	64
90	360	160
# > 0	18,673	7,589

Table 3

Returns to Training for Different Functional Forms, NLSY

Specification	$\bar{R}^2$	Fraction Fourier Series Explained	Total Effect at Median
Complete Sample			
No Training Vars.	0.2033	--	--
Dummy	0.2042	.624	0.031
Linear	0.2040	.332	0.003
Quadratic	0.2042	.461	0.005
Cube root	0.2050	.842	0.036
Log	0.2049	.822	0.041
Dummy + Linear	0.2047	.732	0.029
Fourier series	0.2057	--	0.039
N	17,809		
Obs	75,698		
Training Outliers Omitted*			
No Training Vars.	0.2023	--	--
Dummy	0.2032	.630	0.031
Linear	0.2031	.481	0.007
Quadratic	0.2035	.716	0.014
Cube root	0.2040	.842	0.037
Log	0.2038	.823	0.041
Dummy + Linear	0.2037	.762	0.029
Fourier series	0.2047	--	0.034
N	17,788		
Obs.	75,497		

*Top 1% of training duration omitted.

Table 4

Descriptive Statistics, EOPP

Variable	Mean	Std. Dev.	Min.	Max.
Ln Wage Growth	0.19	0.20	-.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
Hrs. formal tr., formal tr. > 0	72.27	101.14	1.00	640.00
Hrs. informal tr., informal tr. > 0	131.72	175.03	1.00	2070.0
Ln (formal tr. + 1), formal tr. > 0	3.57	1.23	0.69	6.46
Ln (informal tr. + 1), inf. tr. > 0	4.23	1.21	0.69	7.64
Ln # weeks until fully trained	2.21	1.24	0.00	6.033
Years relevant experience	2.38	4.49	0.00	40.00
Rel. experience squared	25.76	108.31	0.00	1600.00
Age	26.89	9.10	16.00	64.00
Years education	12.47	1.65	2.0	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
Obs.	1,715			

Table 5

Returns to Training for Different Functional Forms, EOPP

Specification	$\overline{R}^2$	Fraction Fourier Series Explained	Total Effect at Median
Complete Sample			
No Formal Training Vars.	0.1756	--	--
Dummy	0.1807	.587	0.044
Linear	0.1813	.660	0.014
Quadratic	0.1837	.951	0.031
Cube root	0.1834	.873	0.047
Log	0.1830	.840	0.052
Dummy + Linear	0.1822	.790	0.036
Fourier series	0.1833		0.041
Obs	1,715		
Training Outliers Omitted*			
No Formal Training Vars.	0.1749		--
Dummy	0.1801	.574	0.045
Linear	0.1829	.837	0.018
Quadratic	0.1831	.963	0.028
Cube root	0.1834	.899	0.050
Log	0.1827	.847	0.053
Dummy + Linear	0.1832	.902	0.035
Fourier series	0.1830		0.041
Obs.	1,713		

*Top 1% of formal training duration observations omitted.

Table 6

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 (%)
Complete Sample				
No Training Vars.	0.1949	--	--	--
Dummy	0.1961	.581	.042	140
Linear	0.1957	.299	.004	12
Quadratic	0.1959	.378	.007	23
Cube root	0.1969	.772	.045	149
Log	0.1968	.759	.053	175
Dummy + Linear	0.1965	.683	.040	132
Fourier series	0.1970	--	.059	197
	16,534			
Obs	69,800			
Training Outliers Omitted*				
No Training Vars.	0.1939	--	--	
Dummy	0.1951	.734	.042	140
Linear	0.1951	.590	.008	28
Quadratic	0.1956	.790	.018	59
Cube root	0.1959	.978	.048	159
Log	0.1958	.949	.053	178
Dummy + Linear	0.1957	.914	.040	134
Fourier series	0.1960	--	.056	186
n	16,502			
Obs.	69,573			

*Top 1% of training duration omitted.

Table 7

Fourier Series Estimates of Mean Rates of Return for 25th - 75th Percentiles of Positive Training Distribution

Without correction for heterogeneity in growth rates:	
Complete sample	183 (33)
Outlier-omitted sample	178 (33)
Corrected for heterogeneity in growth rates:	
Complete sample	128 (36)
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)

*Standard errors are in parentheses

Table 8

Rates of Return to Training for Different Functional Forms with Lagged and Lead Training and Final Training Interactions, NLSY

Specification	$\bar{R}^2$	Fraction Fourier Series Explained	Total Effect at Median	Implied Rate of Return at Median (%)
Complete Sample				
No Training Vars.	0.1970	--	--	--
Dummy	0.1973	.411	0.021	69
Linear	0.1970	.063	0.001	2
Quadratic	0.1971	.099	0.001	5
Cube root	0.1974	.518	0.025	82
Log	0.1974	.539	0.029	95
Dummy + Linear	0.1973	.449	0.022	72
Fourier series	0.1986	--	0.043	144
n	16,534			
Obs	69,800			
Training Outliers Omitted*				
No Training Vars.	0.1959	--	--	
Dummy	0.1962	.681	0.021	71
Linear	0.1960	.316	0.003	10
Quadratic	0.1963	.634	0.009	30
Cube root	0.1963	.892	0.029	95
Log	0.1963	.895	0.030	102
Dummy + Linear	0.1963	.832	0.024	81
Fourier series	0.1966	--	0.040	133
n	16,502			
Obs.	69,573			

Table 9

Selected Coefficients and Rates of Return, Cube Root Specification

	Full Sample		Outliers Omitted	
Lead Training$^{1/3}$	0.0024 (0.0014)	0.0006 (0.0014)	0.0027 (0.0018)	0.0010 (0.0018)
Current Training$^{1/3}$	0.0051 (0.0012)	0.0043 (0.0012)	0.0050 (0.0015)	0.0044 (0.0013)
Lagged Training$^{1/3}$	0.0040 (0.0010)	0.0013 (0.0011)	0.0045 (0.0012)	0.0019 (0.0015)
Final Training$^{1/3}$ x Tenure		0.0021 (0.0008)		0.0020 (0.0009)
Final Training$^{1/3}$ x Tenure^2		-0.0001 (0.0001)		-0.0001 (0.0001)
Final Training$^{1/3}$ x Tenure^3/100		0.0001 (0.0005)		0.0001 (0.0005)
Effect of Training at Median Positive Hours	0.0448 (0.0070)	0.0246 (0.0084)	0.0476 (0.0075)	0.0285 (0.0088)
Rate of Return to Training at Median Positive Hours	149	82	159	95

Table 10

Effect of Tenure on Returns to Training, Outlier-omitted Sample

Interactions of Tenure, Tenure^2, Tenure^3 with:	Final Training	Final Training, Current, Lead and Lagged Training	Final Training, Current Training
Effect of Median Positive Training at:			
25th Percentile of Tenure (1.1 years)	0.0285	0.0365	0.0369
	(0.0088)	(0.0175)	(0.0120)
50th Percentile of Tenure (2.5 years)	0.0285	0.0283	0.0266
	(0.0088)	(0.0127)	(0.0101)
75th Percentile of Tenure (5.0 years)	0.0285	0.0226	0.0222
	(0.0088)	(0.0096)	(0.0089)
Slope of Training Effect with Respect to Tenure at:			
25th Percentile of Tenure		-0.0087	-0.0099
		(0.0053)	(0.0042)
50th Percentile of Tenure		-0.0046	-0.0051
		(0.0035)	(0.0027)
75th Percentile of Tenure		0.0007	0.0011
		(0.0027)	(0.0022)
Effect of Median Positive Training on Wage Growth at Median Tenure (Final Training Interactions)	0.006	0.008	0.008
p value, Tenure-Training Interactions*		0.57	0.12

* Current, lead and lagged training in column (2); current training in column (3).

Table 11

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)	
Lagged Training$^{1/3}$	0.0016	
	(0.0013)	
Current Training$^{1/3}$	0.0012	
	(0.0017)	
Initial Occ. Managerial/Professional x Current Training$^{1/3}$	0.0063	
	(0.0013)	
Initial Occ. Other White Collar x Current Training$^{1/3}$	0.0015	
	(0.0011)	
Part-Time x 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 1000 hours.

Table 12

Interactions of Training and Job Complexity, EOPP

Coefficient	Cube Root Formal Training, Log Informal Training	Cube Root Formal Training, Log Informal Training, Formal Training Interaction	Cube Root Formal Training, Log Informal Training , Training Interactions	Perfect Substitutes	Perfect Substitutes, Job Complexity Interaction
Cube Root Formal Training	0.014 (0.004)	0.006 (0.010)	0.008 (0.010)		
Log Informal Training	0.015 (0.003)	0.015 (0.003)	0.008 (0.006)		
Log Number of Weeks Until Qualified	0.012 (0.004)	0.011 (0.004)	-0.000 (0.009)	0.015 (0.003)	
Cube Root Formal x Log Number of Weeks		0.003 (0.004)	0.002 (0.003)	-0.001 (0.003)	
Log Informal x Log Number of Weeks			0.003 (0.002)		
Log Weighted Aggregate Training				0.026 (0.004)	0.013 (0.008)
Log Weighted Aggregate Training x Job Complexity					0.0055 (0.0028)
Weight on Formal Training				0.935 (0.034)	0.928 (0.038)
Obs.	1,715	1,715	1,715	1,715	1,715
$\bar{R}^2$	0.1834	0.1834	0.1841	0.1844	0.1857

Figure 1
Predicted effect of training on ln wages, various specifications

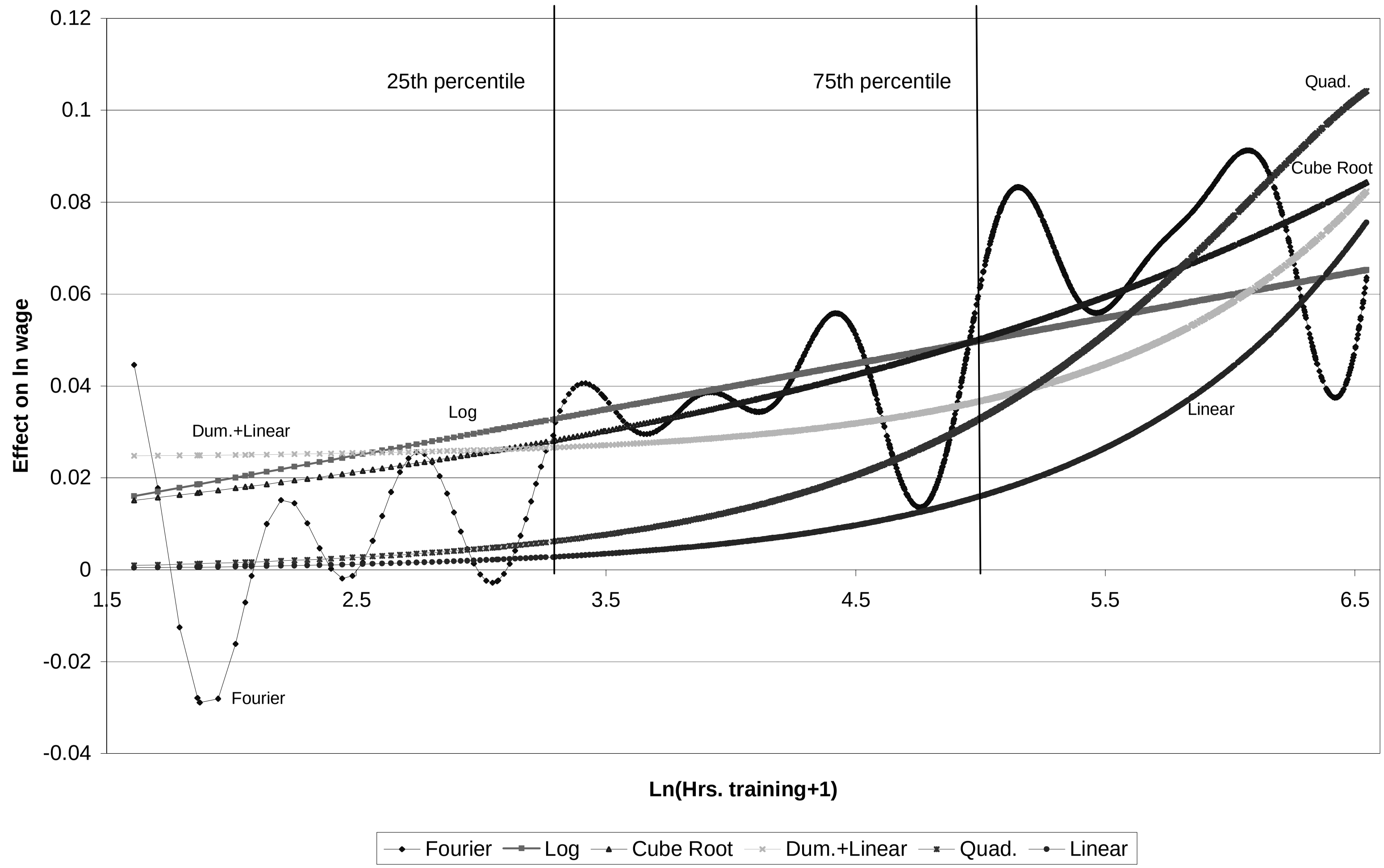
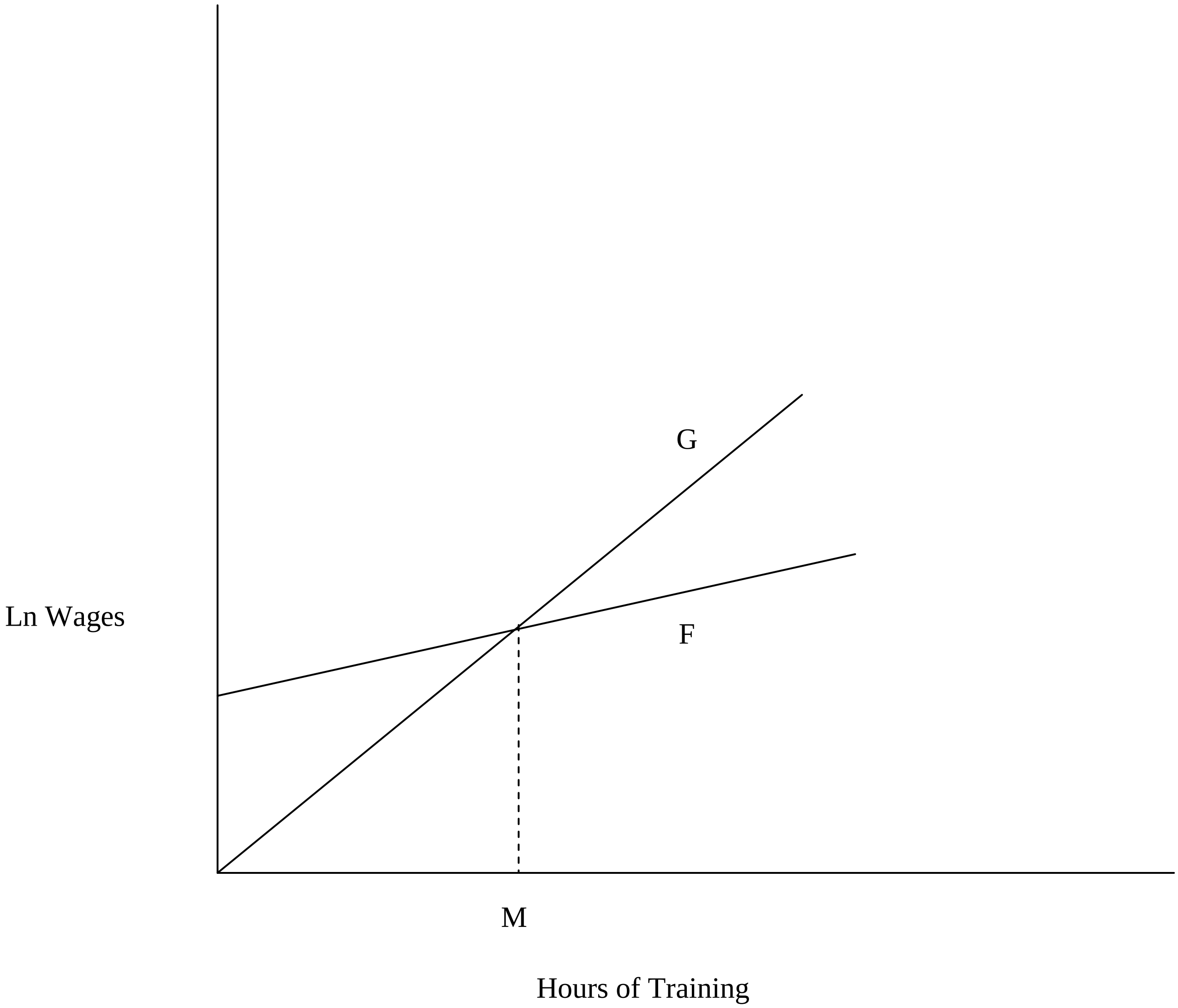


Figure 2



Appendix.

In this appendix, we give the details of our calculation of the maximum effect of duration error and of the potential effects of misclassification error. Let $\eta = \ln(T) - \ln(T^*)$ denote the measurement error in (non-misclassified) log training. We assume that η is independent of $\ln(T^*)$, and is distributed $N(0, \sigma_\eta^2)$ and that $\ln(T^*)$ is distributed $N(\mu, \sigma_*^2)$. These measurement error assumptions are consistent with our data in that the training distributions in the NLSY and EOPP are both approximately log-normal.[33] In fact, note that 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.

Under our assumptions, the distribution of $\ln(T^*)$ conditional on $\ln(T)$ is normal with mean $E(\ln(T^*)\,|\,\ln(T)) = \mu + \frac{\sigma_*^2}{\sigma_T^2}(\ln(T) - \mu)$ and variance

(A1) $\sigma^2_{\ln(T^*)|\ln(T)} = (1-\rho^2)\,\sigma_*^2$,

where $\rho = \frac{\sigma_*}{\sigma_T}$ denotes the correlation coefficient between $\ln(T^*)$ and ln(T), and $\sigma_T^2 \equiv \sigma_*^2 + \sigma_\eta^2$ is the variance of observed log training. Consequently, if the return to training is a linear function of $\ln(T^*)$, say $g(T^*) = \beta\ln(T^*)$, then

(A2) $E(\ln W|\ln(T) = \mu) = \beta\, E(\ln(T^*)|\ln(T) = \mu) = g(\exp(\mu))$,

so that the return to training is consistently estimated at the geometric mean of training.

---

[33] Estimating the Box-Cox transformation to normality $(T^\lambda - 1)/\lambda$ 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 –.01 in the NLSY. Quantile plots also show that log-normality is a good approximation in both datasets.

If the return to training is not a linear function of ln(T^*), then the effects of measurement error will not cancel out at the geometric mean, so that E(ln W|ln(T)= μ) ≠ $g(\exp(\mu))$. To see this, let $h(\cdot)$ be the function implicitly defined by $h(\ln(x)) = g(x)$, and note that E(ln W|ln(T)= μ) = E(h(ln(T^*))|ln(T)= μ) = $\int_{-\infty}^{\infty} h(T^*)\gamma(\ln(T^*))d\ln(T^*)$, where $\gamma(\cdot)$ is the (normal) density function for ln(T^*) conditional on ln(T) = μ. Taking a second-order Taylor expansion of $h(\cdot)$ around μ, the expected return to training for an individual with observed non-misclassified training ln(T) = μ can be expressed as

$$\text{(A3)}\quad E(\ln W|\ln(T)=\mu) = h(\mu) + h'(\mu)\int_{-\infty}^{\infty}(\ln(T^*)-\mu)\gamma(\ln(T^*))d\ln(T^*)$$

$$+ (1/2)\int_{-\infty}^{\infty} h''(\tau(\ln(T^*)))(\ln(T^*)-\mu)^2\gamma(\ln(T^*))d\ln(T^*),$$

$$= g(\exp(\mu)) + (1/2)\int_{-\infty}^{\infty} h''(\tau(\ln(T^*)))(\ln(T^*)-\mu)^2\gamma(\ln(T^*))d\ln(T^*),$$

where $\tau(\ln(T^*))$ is some value between μ and ln(T^*), and the second term on the right hand side of the first equation is 0 by virtue of the fact that E(ln(T^*)|ln(T)= μ. From (A3), we see that the nature of the bias at μ is determined by the sign of h''. If the true return to training declines at a slower than logarithmic rate so that $h'' > 0$, the estimated return to non-misclassified observed training at the geometric mean will exceed the true return.

To estimate the potential upward bias if $h'' > 0$, let the return to true training be given by $g(T^*) = c(T^*)^{\delta}$, $0 < \delta < 1$, which means that $h(x) = c(\exp(x))^{\delta}$. Further expanding the Taylor series in (A3), using the fact that $h^{(n)}(x) = \delta^n h(x)$, and rearranging terms, the proportional upward bias in the estimated return to training at the geometric mean is

$$\frac{E(\ln W \mid \ln(T)=\mu)-h(\mu)}{h(\mu)}=\sum_{n=1}^{\infty}\frac{\delta^{2n}}{(2n)!}E((\ln(T^*)-\mu)^{2n} \mid \ln(T)=\mu).$$

Using (A1) and maximizing $\sigma^2_{\ln(T^*)|\ln(T)=\mu}$ with respect to σ^2_* while holding σ^2_T constant, one can show that $\sigma^2_{\ln(T^*)|\ln(T)=\mu} \leq (1/4)\sigma^2_T$. The variance of observed log training in the NLSY is 1.94. In EOPP, the variance of observed log formal training is 1.27. Numerical calculations thus show that if δ = .33, the maximum proportional bias in the estimated return to geometric mean training is only about 3 percent in the NLSY and 2 percent in EOPP. If δ = .75, the maximum proportional bias in the estimated return to geometric mean training is about 18 percent in the NLSY and 11 percent in EOPP.

Classification error will cause underestimation of the returns at the geometric mean (and all other points). We can make a plausible estimate of the extent of classification error by using data from a 1993 survey matching workers' and firms' reports of training sponsored by the Upjohn Institute, similar in design to EOPP (Barron, Berger, and Black 1997).[34] The Upjohn survey covers formal (and informal) training on the first four weeks on the job.

We first deal with forgotten training, assuming for the time being that there is no false training. Recall that training effects are $g(T) - g(0)$, but we estimate $f(\mathrm{T}) - f(0)$. With forgotten training, $f(0) \equiv \mathrm{E}(\ln W|T=0) = \mathrm{E}(g(T^*)|T = 0)$. Let

$r \equiv \frac{(1-p)E(g(T^*) \mid T=0)}{(1-p)E(g(T^*) \mid T=0)+pE(g(T^*) \mid T>0)}$ denote the proportion of the total $g(T^*)$ that is forgotten. Disregarding duration error, $f(T_0)=g(T_0)$ and

[34] We thank Dan Black for supplying us with the data from this survey.

$E(g(T^*)\,|\,T=0)=\dfrac{prE(g(T^*)\,|\,T>0)}{(1-p)(1-r)}=\dfrac{prE(g(T)\,|\,T>0)}{(1-p)(1-r)}$. Normalizing $g(0)$ to zero, the proportionate bias in the estimated return to training is thus given by

$$\frac{f(T_0)-f(0)-(g(T_0)-g(0))}{g(T_0)-g(0)}=-\frac{prE(g(T)\,|\,T>0)}{(1-p)(1-r)g(T_0)}.$$

We assume that reports of formal training by the employer that are not mentioned by the employee are forgotten training. Taking $g(\cdot)$ to be the cube root, the Upjohn data indicate that about 33 percent of the cube root of formal training is forgotten. In the NLSY, p = .1 and $E(g(T)\,|\,T>0)$ is 4.36. Assuming that the proportion r of the total $g(T^*)$ that is forgotten is the same in Upjohn as in the NLSY, forgotten training depresses the estimated rate of return by 6%.

Now consider the effect of false training. Consider an example where the probability of forgetting a spell of training is .35 (compatible with Upjohn data) and the probability of reporting a spell where none exists is .01, where the observed probability of training is .1 as in the NLSY, and where the distribution of false training is the same as the distribution of true observed training. Note that $p=p^*(1-\alpha_1)+(1-p^*)\alpha_0$, where p^* is the true probability of training, α_0 is the probability of false training (conditional on no true training), and α_1 is the probability of forgetting training conditional on receiving training. The percentage of reported training that is false is $(1-p^*)\alpha_0\,/\,p$, which solving for p^* from the parameter values just given is 8.6%.

In summary, even small amounts of false training cause substantial bias when the probability of training is relatively low. In our example, a false training probability of 1% leads to an 8.6% downward bias in the estimated return to training. Combining this bias with the bias from forgotten training, one obtains a number of the same order of magnitude as the maximum positive bias from duration error, so on net it seems unlikely that there is significant positive bias to estimated returns at the geometric mean.

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 greatly complicates the analysis. The EOPP, with a single formal training item, is not subject to this problem.

One factor that might work in the direction of overestimating the return to training is 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 false training is reported. If, as indicated by Loewenstein and Spletzer (1996), training is highly correlated within jobs and if respondents sometimes forget to report training, the amount of training that respondents with positive training receive on average could be underestimated, resulting in overestimation of returns. On the other hand, the sum of measurement error across multiple spells will tend to become less similar to the log-normal distribution and more similar to the normal, which would tend to push down estimated returns at the geometric mean for any given functional form.

We conducted a small Monte Carlo exercise in order to judge the probable effect of the panel nature of our data on the direction of the measurement error effect. The set-up is as follows. Workers are observed for four periods t={0,1,2,3}. Time 0 corresponds to the start of the job, before training is observed. Training spell duration is positive in periods 1-3 if the latent variable $T^{**} > 0$. $T_{it}^{**} = K - .021t + .0006t^2 + \rho\lambda_i + \sqrt{1-\rho^2}\varepsilon_{it}$, where λ_i is a person-specific fixed effect and ε_{it} is a residual uncorrelated across periods, both distributed normally with unit variance, and where ρ is a parameter that varies across specifications. The coefficients on tenure, here and below, are taken from the NLSY data. Training is forgotten with a constant probability α. The parameter K is a function of α such that the observed probability of training over the four periods is .10, corresponding to the data.

Training duration for positive training observations is distributed log-normally, with $\ln(T_{it}^* \mid T^{**} > 0) = 3.84 - .091t + .0052t^2 + e$, where e is distributed normally with mean zero and is uncorrelated with ε and across periods. In specifications without duration error, e has standard deviation 1.27. In specifications with duration error, e has standard deviation $1.27/\sqrt{2}$ and a normal mean zero duration error variable with the same standard deviation (and independent of the other variables) is added to produce observed log training $\ln(T_{it})$ (when training is not forgotten). Geometric mean training for an individual spell is about 40 hours with this specification, corresponding to the data.

Wages are generated by the following process:

$\ln(W_{it}) = 1.92 + \beta g(\sum_{\tau \le t} T_{i\tau}^*) + .095t + .0021t^2 + \kappa_i + u_{it}$, where κ is a normally-distributed person-specific effect with mean zero and standard deviation 0.5 and u is a normally-distributed residual uncorrelated across periods with standard deviation 0.2. The function g varies across specifications, with the parameter β varying such that 60 hours of training increases log wages by .04.

For computational simplicity, we simulate estimation by splines, taking quartiles of the positive training distribution as our knot-points. We take as our bias measure the percentage bias of the predicted effect of training at the median of positive training for each simulated sample. For comparison purposes, we also simulate comparable specifications where we observe only periods t={0,1}. We computed 1,000 simulations per specification.

Results are shown in Appendix Tables 1 and 2. Table 1 shows the results for the four-period setup. There appears to be a downward bias in each specification due to the spline functional form, in most cases small. (The small reported bias in the linear specification with no duration error and zero probability of forgetting is due to sampling error.) Measurement error in most cases appears to increase the downward bias. The possibility of overestimation due to respondents with positive training sometimes failing to report training appears to be a factor only for extreme specifications--specifically, the linear specification with perfect correlation of the propensity to train across time periods, moderate values of forgotten training, and no duration error. Duration error decreases underestimation in some specifications, but typically in specifications with a great deal of forgotten training, and not enough to lead to overestimation of returns.

Appendix Table 2 shows results from the 2-period setup. These are in line with our theoretical analysis above: for some linear and $T^{.75}$ specifications with no or moderate amounts of forgotten training duration error leads to overestimation of returns. In summary, our Monte Carlo exercise indicates that the fact that the NLSY has more than two periods of data makes it less likely that we are overestimating returns to median hours of training.

Appendix Table 1

Percentage Bias at Median Hours of Training, Four-period Simulations

True functional form	ρ	Probability of forgetting	% bias at median	
			No duration error	Duration error present
Ln	0.4	0	-4.0%	-5.1%
		0.25	-11.9%	-13.7%
		0.50	-29.6%	-28.6%
		0.75	-66.7%	-65.2%
	0.6	0	-4.0%	-4.9%
		0.25	-12.0%	-12.3%
		0.50	-27.5%	-26.8%
		0.75	-62.9%	-62.8%
	0.8	0	-3.2%	-5.3%
		0.25	-12.0%	-12.8%
		0.50	-25.0%	-25.0%
		0.75	-59.8%	-59.7%
	1	0	-7.9%	-10.1%
		0.25	-12.3%	-14.8%
		0.50	-25.5%	-25.8%
		0.75	-52.0%	-50.0%
Cube	0.4	0	-5.6%	-9.1%
root		0.25	-14.2%	-15.3%
		0.50	-32.2%	-33.1%
		0.75	-70.4%	-68.5%
	0.6	0	-4.5%	-10.4%
		0.25	-16.4%	-20.0%
		0.50	-35.2%	-36.4%
		0.75	-71.1%	-69.8%
	0.8	0	-4.7%	-11.1%
		0.25	-20.1%	-25.9%
		0.50	-39.8%	-42.8%
		0.75	-70.5%	-73.0%
	1	0	-34.2%	-53.7%
		0.25	-52.9%	-63.7%
		0.50	-63.9%	-68.8%
		0.75	-80.7%	-81.2%

Appendix Table 1, continued

True functional form	ρ	Probability of forgetting	% bias at median	
			No duration error	Duration error present
$T^{.75}$	0.4	0	-4.4%	-5.0%
		0.25	-14.5%	-11.6%
		0.50	-28.6%	-22.9%
		0.75	-69.4%	-59.1%
	0.6	0	-5.2%	-6.3%
		0.25	-11.4%	-13.7%
		0.50	-27.9%	-22.6%
		0.75	-69.1%	-55.6%
	0.8	0	-3.9%	-10.8%
		0.25	-9.4%	-17.0%
		0.50	-24.2%	-25.2%
		0.75	-64.0%	-55.5%
	1	0	-8.6%	-21.2%
		0.25	-10.8%	-25.9%
		0.50	-21.7%	-26.7%
		0.75	-56.9%	-50.3%
Linear	0.4	0	-1.0%	-2.1%
		0.25	-6.2%	-6.4%
		0.50	-22.7%	-14.6%
		0.75	-75.2%	-47.0%
	0.6	0	2.1%	-7.0%
		0.25	-5.5%	-6.3%
		0.50	-14.2%	-13.3%
		0.75	-67.5%	-43.3%
	0.8	0	-1.1%	-12.1%
		0.25	0.4%	-10.4%
		0.50	-8.8%	-10.4%
		0.75	-59.4%	-40.2%
	1	0	-0.3%	-12.0%
		0.25	14.2%	-3.6%
		0.50	18.3%	1.5%
		0.75	-31.4%	-22.9%

Appendix Table 2

Percentage Bias at Median Hours of Training, Two-period Simulations

True functional form	Probability of forgetting	% bias at median	
		No duration error	Duration error present
Ln	0	-3.8%	-5.0%
	0.25	-6.4%	-9.1%
	0.50	-15.2%	-18.2%
	0.75	-44.6%	-45.4%
Cube root	0	-4.3%	-3.2%
	0.25	-9.2%	-7.5%
	0.50	-18.9%	-17.0%
	0.75	-50.0%	-45.3%
$T^{.75}$	0	-7.6%	5.7%
	0.25	-15.7%	0.2%
	0.50	-28.9%	-10.8%
	0.75	-72.0%	-46.0%
Linear	0	-1.5%	14.2%
	0.25	-11.3%	8.5%
	0.50	-30.6%	-5.2%
	0.75	-91.3%	-45.2%