BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF COMPENSATION AND
AND WORKING CONDITIONS

Belated Training: The Relationship Between Training, Tenure, and Wages

Mark A. Loewenstein and James R. Spletzer

Working Paper 296
September 1996

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. We thank Harley Frazis and seminar participants at Rutgers University, the Census Bureau's Center for Economic Studies, and the NLS Summer Workshop for helpful comments.

Belated Training:

The Relationship Between Training, Tenure, and Wages

September 13, 1996

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

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

James R. Spletzer
Bureau of Labor Statistics
2 Massachusetts Ave NE, Room 4945
Washington D.C. 20212

Phone: 202-606-7393
E-Mail: Spletzer_J@bls.gov

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. We thank Harley Frazis and seminar participants at Rutgers University, the Census Bureau's Center for Economic Studies, and the NLS Summer Workshop for helpful comments.

Abstract

The increasing and concave age earnings profile is one of the most commonly accepted facts in economics. The human capital model attributes rising wage-tenure and wage-experience profiles to investments in worker productivity. Therefore, an implicit assumption in the human capital model is that training must be occurring throughout a worker's stay at an employer. However, the empirical training literature has thus far either implicitly or explicitly assumed that training is concentrated at the beginning of the employment relationship. Using data from the National Longitudinal Survey of Youth, we find that training in the later years of tenure is the norm rather than the exception. To eliminate the bias from belated information, we estimate a fixed effects wage equation with a two stage procedure that uses previous training and fixed individual and match-specific variables as instruments. While instrumental variables estimation is the standard method of correcting for measurement error, instrumental variables estimates are not consistent when the explanatory variable measured with error (in our case, training) can only take on a limited range of values. Nevertheless, the relative magnitudes of the OLS and the IV training coefficients provide information that enables us to correct for measurement error bias without any need for external information on misclassification rates. Our results provide strong support for the human capital model in that most wage growth beyond the first year of tenure is due to training.

I. Introduction

Age-earnings profiles are virtually always increasing and concave. The most commonly accepted explanation, first offered by Becker (1962) and Mincer (1962), stresses workers' human capital investments while on the job. More recently, some economists have argued that the wage returns to tenure and experience may reflect other factors. For example, a positive return to experience can be explained by the fact that individuals who have been in the labor force longer are more likely to have located good job matches.[1] And Lazear (1981) has argued that a positive return to tenure may reflect the fact that employers defer wage compensation in order to offer workers incentives not to shirk.

Only in recent years has information on explicit measures of on-the-job training been collected, making possible direct tests of the human capital hypothesis. Studies using these data find support for the human capital model's prediction that a worker's wage is positively related to past investments in his training.[2] Indeed, Brown (1989) reports that "within-firm wage growth is mainly determined by contemporaneous productivity growth."

If the human capital explanation of the positive wage-tenure relationship is correct, then training should persist well after the employment relationship has begun. However, the literature as a whole pays little attention to the actual timing of training investments. In fact, largely because of data limitations, most studies seem to assume that training is concentrated at the start of the employment relationship. One exception is a recent paper by Bartel (1995), who analyzes the

---

[1] Similarly, in a cross section, positively sloped wage tenure profiles can be explained by the fact that good matches are more likely to survive. In fact, Abraham and Farber (1987) and Altonji and Shakotko (1987) assert that once one controls for the ultimate duration of a job, tenure has no additional effect on wages. However, Topel (1991) argues that these estimates are biased downward and that tenure is positively related to wages even after one controls for the pure job matching effect.

[2] These studies include Altonji and Spletzer (1991), Barron, Berger, and Black (1994a), Barron, Black, and Loewenstein, (1989, 1993), Bartel (1995), Brown (1989), Duncan and Hoffman (1978), Lillard and Tan (1992), Loewenstein and Spletzer (1994, 1996b), Lynch (1992), Mincer (1988), Parsons (1989), Pergamit and Shack-Marquez (1987), and Veum (1995).

1986-90 personnel records of a large manufacturing company. Bartel finds that “although the probability of receiving training and the amount of training are highest for the newly hired employees, an important observation is that experienced employees at this company continue to receive formal training. For example, in 1990, 47% of the individuals who were hired before 1980 received some formal training. This is a significant finding since it indicates that training is not confined to new hires but is an ongoing process at this company.”

There are in fact strong theoretical reasons why one might expect to see significant amounts of training beyond the first year of employment. In a simple stationary world with linear costs, the human capital model predicts that training will be concentrated at the start of the employment relationship. Under these conditions, there is no reason for an employer to delay a training investment beyond a worker's initial period of employment because this merely reduces the time period over which returns will be realized. Of course, as pointed out some time ago by Ben-Porath (1967) and Blinder and Weiss (1976), a convex cost function for training can cause training to be spread out over time.

Perhaps more importantly, the simple argument above implicitly presumes that the training cost function is invariant over time and that the employer and worker both have full information about the quality of their match at the beginning of the employment relationship. For example, it seems quite likely that a worker may be better able to absorb expensive training after an initial period of "learning by doing" in which he has been acclimatized to his job and work environment. If the cost of training the worker is sufficiently lower after he has had some experience with the employer, then it may pay the employer to defer some training until it is cheaper, even if this means foregoing the returns to training during the early part of the employment relationship.

Furthermore, as pointed out some time ago by Johnson (1978) and Jovanovic (1979), information about the quality of a firm-worker match is often revealed only after a period of time. An employer who delays training can target his training investments on his most promising workers, perhaps concentrating on training only those workers who are targeted for promotion. In addition, by deferring training, an employer and worker can significantly lower the probability of making a costly investment in a bad match that is unlikely to last very long (either because the worker turns out to have a low match-specific productivity or because he discovers that he does not like the non-pecuniary attributes of the employer's job).

In this paper, we use recent data from the National Longitudinal Surveys of Youth in order to thoroughly analyze the relationship among training, tenure, and wages. Our incidence analysis in section II indicates that a significant amount of recurring and delayed training occurs well beyond the start of the employment relationship. In section III, we turn to the motivating question of the human capital model's prediction that wage growth is caused by productivity increases and the implication that wage growth in the later years of tenure is at least partly caused by belated spells of training. In order to eliminate potential biases from individual fixed effects, match specific fixed effects, and belated information, we estimate fixed effects wage equations using a two stage procedure that corrects for the endogeneity of training in the wage growth equation. One of our key results is that the bias due to measurement error in training incidence is quite large.

As discussed in section IV, instrumental variable estimation in the presence of measurement error is complicated by the fact that the classical errors in variable model is not applicable when the explanatory variable that is measured with error can only take on a limited range of values. We show that the relative magnitudes of the IV and the OLS training coefficients provide information about the amount of measurement error. This enables us to correct for measurement error bias without relying on external information on misclassification rates such as might be provided by a validation study. After correcting for measurement error bias, we find that the returns to training are quite large and much larger than the tenure coefficients that measure wage growth in the absence of training. Similar to Brown (1989), these results provide strong support for the human capital model.

II. The Relationship Between Training and Tenure

In this section, we examine the relationship between training and tenure using data from the National Longitudinal Survey of Youth (NLSY).[3] The NLSY is a dataset of 12,686 individuals who were aged 14 to 21 in 1979. These youth have been interviewed annually since 1979, and the response rate has been 90 percent or greater in each year. The NLSY data contain detailed information on wages, tenure, and training. We utilize data from the 1988 through 1991 NLSY surveys because these data provide information on all training spells regardless of their duration.[4] To ensure that we observe all training spells at a given employer, we restrict our sample to individuals who start a new job within one year of the 1988 survey or later. This restriction leaves us with between one and four years of data for individuals aged 23-34. Because we omit persons with more than one year of tenure in 1988 who do not change jobs between 1988 and 1991, our sample likely has an above average proportion of high turnover individuals. After eliminating observations where an individual is not employed or where there are missing training data, we end up with a sample made up of 6,145 individuals who in total contribute 15,743 person-year observations. The NLSY training question is "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?" In our sample of 15,743 person-year observations, the average annual training incidence is 12.18 percent.

---

[3] Lynch (1991, 1992) characterizes training incidence and determines the effects of training on wages and turnover using the 1979-1983 NLSY data. Loewenstein and Spletzer (1996b) use the 1988-1991 NLSY data to analyze the division of the costs and returns to training. Veum (1995) also analyzes the 1988-1991 NLSY data.

[4] The training section of the NLSY questionnaire was re-designed in 1988. Prior to 1988, information was only obtained on training spells that lasted longer than one month. Data from 1988-1991 indicate that 64.5% of training spells are less than four weeks in duration. There were no training questions in the 1987 survey. Data from 1992 and 1993 were not yet available at the time we started this project.

Table 1 provides information on the relationship between a worker's tenure and the probability that he receives training in a given year. The integer value of tenure=1 in column 1 applies to those persons with tenure between 1 week and 52 weeks, tenure=2 applies to those persons with tenure between 53 and 104 weeks, and so forth. The statistics in column 2 are the sample sizes associated with each year of tenure, and the statistics in column 3 are the probabilities that individuals with the specified tenure received training on the current job in the previous year.

The training incidence statistics appearing in table 1 indicate quite clearly that a substantial amount of training occurs after the first year of employment. As can be seen in column 3, annual training incidence nearly doubles from 9.89 percent in the first year of tenure to 17.21 percent in the second year of tenure, and is then roughly constant around 18 percent in the third and fourth years of tenure.[5]

A question that comes immediately to mind after looking at table 1 is the extent to which the training spells occurring after the first year of tenure are first time delayed training spells and the extent to which they constitute recurring training spells by individuals who have already received training in the past. Altonji and Spletzer (1991) find that previous training is associated with a higher probability of present training, but their data does not enable them to determine whether or not this previous training occurred in the current job. Lynch (1992) and Mincer (1988) show that individuals who have received training in past jobs are more likely to receive training in their present job, but they do not look at the relationship between the likelihood of current training and previous training in the current job. To determine the empirical relationship between current

[5] Recall that we are classifying a person as having one year of tenure if he has been on his current job between 1 and 52 weeks at the time of his interview. Our estimate of first year training may therefore understate the total amount of training that takes place in the first 52 weeks of employment. Although this alone might lead us to overstate the amount of training provided in the second year of tenure, we are likewise misclassifying training in the latter part of the second year of tenure as occurring in the third year, and so forth. Note too that while our estimate of first year training may understate the total amount of training that takes place in the first 52 weeks of employment, it is an overestimate of training that takes place at the very start of the match.

training and previous training in both the current job and in past jobs, we decompose the belated training spells in our sample into those that are first time delayed spells and those that are recurring spells.

Before turning to the data, we should note that the relationship between the receipt of previous training and the probability of current training is theoretically ambiguous, as there are two conflicting effects. Diminishing returns to training by themselves would lead to an *inverse* relation because they would imply that, other things the same, the expected gain to providing current training is lower for individuals who have had previous training than for individuals who have not had previous training. Of course, other things may not be the same. There is considerable heterogeneity among workers. Some can absorb training more easily and cheaply than others. The existence or lack of previous training provides a signal about the individual. To the extent that previous training signals a comparative advantage in the receipt of training, we would expect individuals who have received previous training to be *more likely* to receive current training.

The preceding argument focuses on characteristics that are specific to individuals and thus applies to both previous training in the current job and training in a past job. However, job matching considerations suggest that the probability of current training may be more closely related to previous training in the current job than to previous training in past jobs. Since it only pays to incur the costs of specific training if the match between a worker and an employer is a good one that is likely to last a long time, the occurrence of training indicates that a match is high quality. If training has a larger return in high quality matches, individuals who have received previous training in the current job may be more likely to receive a recurring spell of training in the current year.

Let us now estimate the effect of previous training on the probability of current training. To help fix ideas, let T_τ be an indicator variable equal to one if a worker in his τ^{th} year of tenure receives training during year τ, and let PCJ_τ be the number of previous spells of training that a worker in his τ^{th} year of tenure has received in his current job (only counting one spell per year). The probability of receiving training during the τ^{th} year of tenure can be written as

$$(1) \qquad \Pr(T_\tau = 1) = \sum_{k=0}^{\tau-1} \Pr(T_\tau = 1 | PCJ_\tau = k) \Pr(PCJ_\tau = k) ,$$

where $\Pr(T_\tau=1|PCJ_\tau=k)$ is the worker's probability of receiving training in his τ^{th} year of tenure conditional on his having received k previous spells of training in his current job. Equation (1) indicates that the probability of receiving training during the τ^{th} year of tenure is merely the weighted sum of the conditional probabilities $\Pr(T_\tau=1|PCJ_\tau=k)$, where the weights are the probabilities of having received k spells of previous training in the current job.

In the absence of explanatory variables, the maximum likelihood estimator of the conditional probability $\Pr(T_\tau=1|PCJ_\tau=k)$ is $\rho_{\tau,k}$, where $\rho_{\tau,k}$ denotes the proportion of individuals with tenure τ and k spells of previous training in the current job who receive training in the current year. This estimator is asymptotically normally distributed with mean $\Pr(T_\tau=1|PCJ_\tau=k)$ and variance $\rho_{\tau,k}(1-\rho_{\tau,k})/N_{\tau,k}$, where $N_{\tau,k}$ denotes the number of sample observations where a worker is in his τ^{th} year of tenure and has received k spells of previous training in the current job.

The maximum likelihood estimates of the conditional probabilities $\Pr(T_\tau=1|PCJ_\tau=k)$ appear in column 1 of table 2. The number of sample observations $N_{\tau,k}$ are also listed in table 2 in the column labeled Cell Size.[6] The results in column 1 indicate quite strongly that individuals with

---

[6] A careful examination of table 2 reveals that the cell sizes sum to less than 15743. To ensure that we do not double count a single training spell that is mentioned at two consecutive interviews, we distinguish between training spells that were ongoing at the time of last year's interview and those that were completed when estimating $\Pr(T_\tau=1|PCJ_\tau=k>0)$. As expected, the estimated

previous training in their current job are more likely to receive current training. For example, the probability of receiving training in the second year of tenure is 13.44 percent if an individual did not receive training during his first year of tenure, but is 30.61 percent if he did receive training during his first year of tenure. The difference between these two probabilities is statistically different from zero.[7] Similarly, individuals are significantly more likely to receive training in their third and fourth year of tenure if they have received previous training at the same job. Furthermore, there is some evidence that the probability of receiving training is positively related to the number of previous training spells. For example, 28.07 percent of individuals in their third year of tenure who have received training in one of the last two years receive training, whereas 45.10 percent of individuals in their third year of tenure who have received training in each of the last two years receive training. This finding that individuals who have received previous training in their current job are significantly more likely to receive training in the current year suggests that there may be substantial individual and/or match fixed effects in training incidence.

While individuals with previous training in their current job are more likely to receive current training than those who have not, the incidence of first time (delayed) training occurring in the later years of tenure is not insignificant. As noted above, the probability of receiving training in the second year of tenure is 13.44 percent if an individual did not receive prior current job training. In years three and four, this probability is still 11 to 12 percent. One possible explanation for this delayed training is belated information about match quality: a worker and his employer may want to defer making significant investments in training until they have had sufficient time to ascertain the quality of their match. A second explanation is that a worker may be better able to absorb

---

probabilities for the former are near one. The estimated probabilities in table 2 are for individuals who do not have an ongoing training spell at the time of the previous interview.

[7] Under the null hypothesis that $\Pr(T_\tau=1|PCJ_\tau=k)=\Pr(T_\tau=1|PCJ_\tau=k-1)$, the test statistic $(\rho_{\tau,k}-\rho_{\tau,k-1})/[\rho_{\tau,k}(1-\rho_{\tau,k})/N_{\tau,k}+\rho_{\tau,k-1}(1-\rho_{\tau,k-1})/N_{\tau,k-1}]^{1/2}$ has a standard normal distribution. For $\tau=2$ and $k=1$, this test statistic is 6.68.

expensive training after an initial period of learning by doing in which he has become acclimatized to his job and work environment.[8]

We can further decompose the sample mean of training for each year of tenure conditional on the number of previous spells of training in the same job into the means for distinct subgroups defined by whether or not a worker has received previous training in a past job. That is, let POJ_τ be an indicator variable equal to 1 if a worker in his τ^{th} year of tenure received previous training in other jobs.[9] Let $Pr(T_\tau=1|PCJ_\tau=k,POJ_\tau>0)$ denote a worker's probability of receiving training in his τ^{th} year of tenure conditional on his having received k previous spells of training in the current job and at least one spell of training in past jobs, and let $Pr(T_\tau=1|PCJ_\tau=k,POJ_\tau=0)$ denote a worker's probability of receiving training in his τ^{th} year of tenure conditional on his having received k previous spells of training in his current job but no spells of training in past jobs. The probability of receiving training during the τ^{th} year of tenure can be written as

$$(2) \qquad Pr(T_\tau = 1) = \sum_{k=0}^{\tau-1} Pr(T_\tau = 1 | PSJ_\tau = k, POJ_\tau = 0)\, Pr(PSJ_\tau = k, POJ_\tau = 0)$$

$$+ \sum_{k=0}^{\tau-1} Pr(T_\tau = 1 | PSJ_\tau = k, POJ_\tau > 0)\, Pr(PSJ_\tau = k, POJ_\tau > 0)\ .$$

---

[8] Technological change is another possible reason for delayed training. Indeed, Bartel and Sicherman (1995) find that technological change increases the incidence of training for individuals who did not receive training in the prior year. For a thorough analysis of delayed training using several different data sets, see Loewenstein and Spletzer (1996a).

[9] Note that we are using an indicator variable rather than the number of previous spells of training in past jobs. Our original specifications controlled for both the number of previous training spells in past jobs and the number of years observed in past jobs. The number of years observed in past jobs controls for the left censoring of our panel. We are able to reject the hypothesis that the training probabilities differ by the number of years observed in past jobs. Although we can not reject the hypothesis that the training probabilities differ by the number of spells observed in past jobs, this appears to be the result of sample size rather than meaningful coefficient differences. Therefore, in order to minimize the number of reported coefficients in table 2, we aggregate the

Maximum likelihood estimates of the conditional probabilities in equation (2) are reported in column 2 of table 2.[10] For individuals in their first year of tenure, previous training in other jobs is associated with a significantly higher probability of training. Specifically, while the probability of receiving training is 8.94 percent for individuals who have not had previous training in past jobs, it is 28.76 percent for individuals who have had such training. However, for other years of tenure, the relationship between previous training in past jobs and current training is not so clear-cut. In fact, while previous training in other jobs tends to raise the probability of training if an individual has had previous training in the current job, it appears to slightly lower the probability of training if the individual has not had previous training in the current job.

The estimates in columns 1 and 2 of table 2 provide some evidence of individual fixed effects and substantial evidence of match fixed effects in training. To determine whether these effects can be explained by observable heterogeneity, we estimate an ordinary least squares equation that includes various individual and job characteristics in addition to variables indicating the presence or absence of prior training. The estimated equation is reported in column 3 of table 2.[11] Although not listed, the coefficients on education and the armed force qualifying test are positive, consistent with both the existing literature and our priors that more able individuals are more likely to receive training. As is apparent from column 3, much of the fixed effect in training cannot be explained by observables. The point estimates of the conditional probabilities fall by

---

number of training spells in past other jobs into an incidence measure. This aggregation does not affect the main conclusions cited in the text.

[10] A careful examination of the cell sizes in table 2 reveals that many individuals in our sample are not observed in past jobs. For example, no past job is observed for 71.4 percent of the observations with tenure=1. We have estimated the probability of receiving training during the τ^{th} year of tenure for those individuals with no observed past jobs, but we do not report these coefficients in table 2.

[11] As is well-known, OLS is not entirely appropriate for estimating a conditional probability model such as this. We present OLS estimates in column 3 because they facilitate comparison with the results in column 2 (in the absence of explanatory variables, the point estimates of the coefficients would be identical). Our conclusions do not change when we estimate the equation in column 3 using a probit model.

approximately 20 to 25 percent when we include the explanatory variables, but most remain statistically different from zero. Furthermore, the pattern that the probability of training increases with the number of previous training spells in the current job still holds when we control for observable heterogeneity.

III. Training and Wage Growth

Our results in the previous section indicate that a substantial amount of recurring and delayed training occurs well beyond the start of the employment relationship. In this section, we investigate the extent to which this belated training can explain wage growth. A finding that training is positively associated with wage growth would be strong evidence in favor of the human capital model. However, some care needs to be taken to ensure that our training coefficients are not biased. Our results in the previous section indicate that there are significant individual and match specific fixed effects in training incidence. Furthermore, it appears that the match specific effect in training incidence may evolve over time as belated information becomes available. If there are similar effects in the wage equation, then simple OLS coefficients will be biased.

To fix ideas, suppose the log wage of a worker i who is in his t^{th} year of tenure at employer j is given by

$$(3) \qquad w_{ijt} = \gamma_t + \sum_{\tau=1}^{t} \beta_\tau T_{ij\tau} + \delta \sum_{k=1}^{j-1} \tilde{T}_{ik} + c_1 X_i + c_2 X_{ij} + c_3 X_{ijt} + u_i + v_{ijt} + \varepsilon_{ijt},$$

where γ_t is a tenure specific effect on wages common to all individuals, $T_{ij\tau}$ is a variable indicating whether individual i received any training in his τ^{th} year of tenure at his current job j, $\tilde{T}_{ik} \equiv \sum_\tau T_{ik\tau}$ is the total number of times that individual i received training at his k^{th} employer, X_i is a vector consisting of individual-specific variables (such as ability, race, or gender), X_{ij} is a vector of match-specific variables (such as employer size), and X_{ijt} is a vector of observable variables that can vary within a match over time (such as the individual's marital status). Using a standard decomposition of the error term, u_i is an individual fixed effect that captures the net wage effect of unobserved person-specific variables, v_{ijt} is an individual-job match effect, and ε_{ijt} is a transitory mean zero error component that is uncorrelated with both the explanatory variables and the fixed effects. Note that equation (3) allows training in different years of tenure to have different wage effects. It also allows for the possibility that training in the current and previous jobs may have different wage effects.[12]

As noted above, we are concerned with the possibility that the individual-job match effect varies over time due to the arrival of belated information. If we assume that the belated information is additive, the individual-job match fixed effect can be notationally written as

$v_{ijt} = v_{ijt-1} + \theta_{ijt} = v_{ij} + \sum_{\tau=1}^{t} \theta_{ij\tau}$, where v_{ij} is the net wage effect of unobserved job match variables at the start of the job and $\theta_{ij\tau}$ is the net wage effect of belated information that arrives during the τ^{th} year of tenure. Equation (3) can thus be rewritten as

$$(4) \qquad w_{ijt} = \gamma_t + \sum_{\tau=1}^{t} \beta_\tau T_{ij\tau} + \delta \sum_{k=1}^{j-1} \tilde{T}_{ik} + c_1 X_i + c_2 X_{ij} + c_3 X_{ijt} + u_i + v_{ij} + \sum_{\tau=1}^{t} \theta_{ij\tau} + \varepsilon_{ijt} \, .$$

Consistent with the findings of others, our results in the preceding section indicate that more educated workers and more able workers are more likely to receive training. Since our AFQT measure is certainly not a perfect measure of ability and since we have no measure of the quality of the job match, there are likely to be significant individual and match specific fixed effects in our

---

[12] For a detailed analysis of the differential wage effects of training in the current and previous

wage equation. If training incidence is positively correlated with these effects (that is, if $E[T_{ijt}\cdot(u_i+v_{ij})]>0$), OLS estimation of (4) will result in training coefficients that are biased upward. We can eliminate this bias by estimating a wage growth equation for individuals who do not change jobs. Specifically, if we take first differences of equation (4) for persons who have been at their current employer for more than one year, we obtain

$$(5) \quad w_{ijt} - w_{ijt-1} = (\gamma_t - \gamma_{t-1}) + \beta_t T_{ijt} + c_3(X_{ijt} - X_{ijt-1}) + \theta_{ijt} + \varepsilon_{ijt} - \varepsilon_{ijt-1}.$$

The positive correlation between T_{ijt} and (u_i+v_{ij}) does not cause any problems for OLS estimation of equation (5) because the error components u_i and v_{ij} are first differenced away.[13] However, while first differencing eliminates the bias caused by the error components u_i and v_{ij}, there still remains the potential bias caused by the belated information effect θ_{ijt}. A positive correlation between training T_{ijt} and belated information θ_{ijt} will cause an upward bias in the training coefficient β_t. As is well known, we can eliminate this bias if we replace the training variable T_{ijt} by an appropriate instrumented value that is correlated with T_{ijt} but not with the belated information θ_{ijt}. While it is often difficult to find suitable instruments, this is not a problem in the current case. Note that previous training does not belong in the fixed effects wage growth equation (5). Since the results in the previous section indicate that previous training is correlated with current training, previous training can be used in the construction of an instrument for current

job, see Loewenstein and Spletzer (1996b).

[13] Note that first differencing the wages of an individual who switches jobs eliminates individual fixed effects but not match specific fixed effects. It is for this reason that we focus on the wage growth of job stayers.

training. The variables X_i and X_{ij} also do not belong in the fixed effects wage growth equation and thus provide another source of exogenous variation in our training instrument.[14]

The instrumental variable technique used above to eliminate the belated information bias is also the standard method of correcting for measurement error. Recent research by Ashenfelter and Krueger (1994) and Card (1995) suggests that measurement error in education may cause the estimated returns to education to be biased downward by as much as 15%. One suspects that, if anything, training is more difficult to measure than education. Indeed, using a unique survey of matched employer-employee responses to the same training questions, Barron, Berger, and Black (1994b) find that the correlation between worker reported and employer reported incidence of on-site (off-site) formal training is only .318 (.377).[15] While a positive correlation between training and belated information would cause an upward bias in the training coefficient in equation (5), measurement error in the training variable would lead to a downward bias. We do not have an a-priori prediction regarding the net effect of these two biases; it is an empirical question whether the instrumented training coefficient will be higher or lower than the OLS coefficient.

Let us now turn to the actual estimation results. Table 3 presents the results of estimating a wage level equation. To facilitate comparison with the results to follow we only include job stayers in the estimation (that is, we only include those observations where the length of tenure is greater than one year). As seen in column 2, when no other explanatory variables are included in the

---

[14] Theoretically, it does not matter if we obtain predicted training incidence from OLS or probit estimation because in either case the predicted training measure will be correlated with the actual measure of training and uncorrelated with the error; see, for example, Kelejian (1971). (The preceding statement assumes that training is measured without error. The case of measurement error is analyzed in detail below.) As is well known, when predicted training is obtained via OLS, two stage least squares and instrumental variable estimation are equivalent. However, as noted by Heckman (1978), the instrumental variable estimator is not equivalent to the two stage estimator when the first stage estimation is probit because residuals from the prediction of training will not generally be orthogonal to the explanatory variables in the wage growth equation. Heckman recommends two stage least squares on the grounds that it is easier to use. Consequently, we choose to report results based on the linear estimator.

equation, the estimated returns to training are quite high - the coefficients on training in the current job range from .1246 to .1993, and the coefficient on training in prior jobs is .1203. As noted above, there is considerable evidence that more able workers are more likely to receive training. To the extent that explanatory variables such as education and AFQT score control for productivity differences among workers we should expect their inclusion in the wage equation to significantly reduce the estimated return to training. This is indeed the case. However, as seen in column 3 of table 3, even when we include a rich set of additional explanatory variables, the estimated returns to training are still substantial, as the coefficients on training in the current job range from .0554 to .0773, and the coefficient on training in prior jobs is .0576.[16]

As discussed above, training coefficients in a simple OLS wage level equation suffer from at least two distinct sources of bias. The fact that high ability workers in high quality job matches are more likely to receive training causes the OLS training coefficients to be biased upward, and the effects of belated information relating to the match specific effect also suggests an upward bias. We can eliminate the bias caused by unobserved fixed individual and match differences by estimating a (fixed effects) wage growth equation. Estimated wage growth equations are reported in table 4. Column 1 of table 4 indicates that when we regress wage growth against tenure indicators alone, all three of the tenure coefficients are positive and the first two are statistically different from zero. The coefficient of .0491 on the indicator variable for the second year of tenure tells us that between the first and second year of tenure workers' wages on average rise by about five percent. Similarly, the coefficient of .0259 on the indicator variable for the third year of

---

[15] In contrast, Ashenfelter and Krueger (1994) report that the correlation for self reported years of schooling between identical twins is .658.

[16] We have reestimated all of our wage equations with the duration of training included as an additional explanatory variable. Once one controls for training incidence, the duration coefficients are small in magnitude and statistically insignificant. This may be due to the fact that duration is usually relatively short (as detailed by Loewenstein and Spletzer (1994), the mean duration of completed formal training spells is 111 hours, but yet the median duration of completed formal training spells is 32 hours) and is likely measured with considerable error.

tenure indicates that between the second and third year of tenure wages grow by over two and a half percent, while the coefficient of .0204 on the indicator variable for the fourth year of tenure indicates that wages grow by about two percent between the third and fourth years of tenure. As seen in column 2, the tenure coefficients fall substantially when training indicators are added to the wage growth equation. In column 3, the addition of (first-differenced) explanatory variables further reduces the tenure coefficients, but has little effect on the training coefficients. In accord with our theoretical discussion, the fixed effects training coefficients are smaller than the training coefficients in the wage level equation.[17] Nevertheless, all the training coefficients in the fixed effects equation are positive and the estimated returns to second and third year training are 4.8 and 6 percent, respectively. The coefficient on third year training is statistically different from zero at the 5 percent level.

Estimating a wage growth equation eliminates the biases that are due to unobserved fixed effects in the wage level equation. However, as discussed above, belated information about match quality could still lead to upward biased training coefficients in a wage growth equation while measurement error could lead to coefficients that are biased downward. To examine these questions further, we estimate the wage growth equation using two stage least squares. Besides previous training in the current job and previous training in other jobs, the explanatory variables in the first stage OLS equation also include fixed demographic and job characteristics. The estimated second stage wage growth equation is reported in column 4 of table 4.[18] The coefficient estimates on second, third, and fourth year training are .063, .1285, and .0814 respectively. These estimated returns to training are markedly higher than those in column 3, suggesting that the downward bias

---

[17] Furthermore, when one uses job movers to estimates the returns to first year training, one finds that the estimated returns to training in the first year are considerably higher than the estimated returns to training in later years. This is consistent with our hypothesis that the OLS training coefficients are partly picking up the returns to good job matches - returns which can be differenced out for job stayers, but not for job movers.

in the OLS training coefficients due to measurement error swamps any possible upward bias stemming from belated information about match quality. Still, it seems hard to believe that the measurement error effect can be nearly as large as our estimated equations suggest. We address this question in the next section.

IV. The Effects of Measurement Error on the Returns to Training

The potential upward bias in simple OLS training coefficients stemming from the endogeneity of training has been the object of a substantial amount of attention, but the downward bias from measurement error has received little, if any, discussion. A comparison of columns 3 and 4 in table 4 reveals that the instrumental variable coefficients on second and third year training are approximately three times higher than the corresponding OLS estimates. And the estimated return to fourth year training is 23 percent higher using the instrumental variable estimation.

Is it reasonable that measurement error can have such a large impact on our estimation results? Exploration of this question requires some additional notation. Let T_{it}* be an indicator variable equal to 1 if individual i receives training during his t^{th} year of tenure (for the rest of this section, we drop the "j" subscript since our wage growth equation is estimated only for job stayers). The econometrician observes reported training T_{it} rather than actual training T_{it}*, where the difference is the measurement error $\omega_{it}=T_{it}-T_{it}*$. Taking this measurement error into account, the wage growth regression becomes:

(6) $$\Delta w_{it} = (\gamma_t - \gamma_{t-1}) + \beta_t T_{it} + c_3 \Delta X_{it} + \varphi_{it}\,,$$

where $\varphi_{it} = \theta_{it} + \Delta\varepsilon_{it} - \beta_t\omega_{it}$

---

[18] The first stage equations for second, third, and fourth year training have R^2s of .16, .19, and .28,

To simplify the exposition, we will assume for now that training is the only explanatory variable affecting wage growth. We will show later that the analysis can be generalized to include other explanatory variables. When $c_3 = 0$, the wage growth equation is given by

$$(6')\qquad \Delta w_{it} = (\gamma_t - \gamma_{t-1}) + \beta_t T_{it} + \varphi_{it}.$$

In this case, it is straightforward to show that the estimated OLS coefficient on year t training will have the probability limit

$$(7)\qquad \operatorname{plim} \hat{\beta}_t^{OLS} = \beta_t - \beta_t \frac{\operatorname{cov}(T_{it}, \omega_{it})}{\operatorname{var}(T_{it})} + \frac{\operatorname{cov}(T_{it}, \theta_{it})}{\operatorname{var}(T_{it})}.$$

In accordance with the discussion in Section III, equation (7) suggests two possible sources of bias in the OLS training coefficients in column 3 of table 4: a non-zero correlation between training T_t and the measurement error ω_t, and a non-zero correlation between training T_{it} and the belated information θ_{it}. While the bias due to belated information, $\operatorname{cov}(T_{it},\theta_{it})/\operatorname{var}(T_{it})$, is presumably positive, we show below that the bias due to measurement error, $-\beta_t \operatorname{cov}(T_{it},\omega_{it})/\operatorname{var}(T_{it})$, is unambiguously negative.[19] A priori, the overall bias in the estimated OLS coefficient is thus of indeterminate sign. The OLS coefficient is biased downward (upward) if the bias due to measurement error is larger (smaller) in magnitude than the bias due to the belated information about match quality.

For notational convenience, define $\eta_t \equiv \Pr(T_{it}^*=1|T_{it}=0)$ as the probability that an individual who reports not receiving training actually receives training, and define $\nu_t \equiv \Pr(T_{it}^*=0|T_{it}=1)$ as the

---

so the correlations between our training measures and training instruments are relatively high.

probability that an individual who reports receiving training does not actually receive training. In the discussion that follows, it will also be helpful to define the conditional probabilities $\alpha_{0t} \equiv \Pr(T_{it}=1|T_{it}^*=0)$ and $\alpha_{1t} \equiv \Pr(T_{it}=0|T_{it}^*=1)$. Note that α_{0t} is the probability that an individual who does not receive training incorrectly reports that he receives training while α_{1t} is the probability that an individual who receives training incorrectly reports that he does not receive training. Furthermore, define $p_t \equiv \Pr(T_{it}=1)$ as the probability of a reported training spell and define $p_t^* \equiv \Pr(T_{it}^*=1)$ as the probability of a true training spell. The joint distribution of T_{it} and T_{it}^* is summarized by the following frequency table:

Row % Column %	$T_{it}^*=0$	$T_{it}^*=1$
$T_{it}=0$	$(1-\eta_t)(1-p_t)$ $(1-\alpha_{0t})(1-p_t^*)$	$\eta_t(1-p_t)$ $\alpha_{1t}p_t^*$
$T_{it}=1$	$\nu_t p_t$ $\alpha_{0t}(1-p_t^*)$	$(1-\nu_t)p_t$ $(1-\alpha_{1t})p_t^*$

In interpreting the above frequency table, note that the probability that an individual neither receives nor reports training can be expressed as either $(1-\eta_t)(1-p_t)$ or $(1-\alpha_{0t})(1-p_t^*)$. A similar statement applies with respect to the other joint probabilities.

Since both training T_{it} and the measurement error ω_{it} are discrete variables, equation (7) can be simplified as:

$$(8) \qquad \operatorname{plim} \hat{\beta}_t^{OLS} = \left(1 - \left[\frac{\Pr(T_{it}=1, T_{it}^*=0) - \Pr(T_{it}=1)\{\Pr(T_{it}=1, T_{it}^*=0) - \Pr(T_{it}=0, T_{it}^*=1)\}}{\Pr(T_{it}=1)\{1-\Pr(T_{it}=1)\}}\right]\right)\beta_t$$

19 More precisely, measurement error lowers the absolute value of the estimated coefficient. The statement in the text presumes that $\beta_t > 0$.

$$+\left[\frac{\mathrm{cov}(T_{it},\theta_{it})}{\mathrm{var}(T_{it})}\right]$$

$$=\left[1-(\nu_t+\eta_t)+b_t\right]\beta_t\,,$$

where $b_t \equiv \left[\frac{\mathrm{cov}(T_{it},\theta_{it})}{\mathrm{var}(T_{it})}\right]\left(\frac{1}{\beta_t}\right)$ will be referred to as the proportional belated information bias.

The proportionate bias of $-(\nu_t+\eta_t)$ due to measurement error in a bivariate variable such as training incidence has been derived previously by Aigner (1973) and Freeman (1984). This bias is unambiguously negative in sign; measurement error in the training variable causes the OLS training coefficient to be biased downward.[20]

Now let us turn to the instrumental variable estimation. The training instrument Z_t has been constructed so as to be uncorrelated with the belated information θ_t. We have not yet made any assumptions regarding the correlation between the instrument Z_t and the measurement error ω_t. In the absence of explanatory variables, the estimated instrumental variables (IV) coefficient on training in the t^{th} year of tenure in equation (6') has probability limit:

---

[20] Using the above frequency table, we can rewrite the OLS bias attributable to measurement error in terms of $\{\alpha_0,\alpha_1,p^*\}$ rather than in terms of $\{\eta,\nu\}$. Holding constant α_0 and α_1, one can easily show that the OLS bias $\eta+\nu$ is a U-shaped function of p* that equals 1 when p* is either 0 or 1. This implies that the OLS bias is especially large when p* is either very small or very large. If p* is close to zero, the signal to noise ratio in reported training is quite small for given levels of α_0 and α_1, which merely says that a large proportion of those reporting training will not actually have received training. Although η will be close to zero, ν will be close to one and the OLS coefficient will be near zero. (Similarly, if p* is close to one, then η will be close to one and ν will be close to zero.) As can be seen from Table 1, only about 18 percent of the individuals in our sample report training in their second, third, and four years of tenure. As a consequence, even if the probabilities of correct responses from individuals who receive and do not receive training are as high as 95 percent (that is, if $\alpha_0 = \alpha_1 = .05$), the OLS training coefficient is biased downward by 25 percent. Furthermore, it is easy to show that the OLS bias $\eta+\nu$ rises with an increase in the probability of an incorrect response. For instance, if the probability that individuals respond incorrectly increases

$$(9) \qquad \text{plim } \hat{\beta}_t^{IV} = \left(1 - \left[\frac{\text{cov}(Z_{it}, \omega_{it})}{\text{cov}(Z_{it}, T_{it})}\right]\right)\beta_t \,.$$

Equation (9) indicates that the IV training coefficient is consistent if and only if the instrument Z_t is uncorrelated with the measurement error ω_t. To determine possible causes of correlation, note that we can always write the predicted training measure as

$$(10) \qquad Z_{it}=\kappa_t+T_{it}^*\delta_t+e_{it},$$

where $\kappa_t=E(Z_{it}|T_{it}^*=0)$ and $\delta_t=E(Z_{it}|T_{it}^*=1)-E(Z_{it}|T_{it}^*=0)$ can be interpreted as the intercept and slope from a regression of the predicted training measure on the true unobserved training measure. Substituting (10) into equation (9) and using the fact that $T_{it}=T_{it}^*+\omega_{it}$, we obtain:

$$(11) \qquad \text{plim } \hat{\beta}_t^{IV} = \left(1 - \left[\frac{\delta_t \text{ cov}(T_{it}^*, \omega_{it}) + \text{cov}(e_{it}, \omega_{it})}{\delta_t \text{ var}(T_{it}^*) + \delta_t \text{ cov}(T_{it}^*, \omega_{it}) + \text{cov}(e_{it}, \omega_{it})}\right]\right)\beta_t \,.$$

Equation (11) indicates two potential sources of bias in the instrumental variables training coefficient. The first of these is a possible correlation between e_t and ω_t. Recall that previous training, fixed individual characteristics, and fixed job match characteristics were the identifying variables used in forming the predicted training variables used to estimate the second stage wage growth equation in column 4 of table 4. Our motivation for using an individual's past training in the construction of an instrument for current training is our earlier finding that individuals who report training in the past are more likely to report training today. Of course, the implicit

from 5 to 10 percent, the bias in the OLS training coefficient escalates from 25 percent to 51 percent.

assumption here is that the positive correlation between past and present training arises at least in part because true training is correlated over time. However, the positive correlation between past and present training could occur if individuals tend to consistently misreport training over time. If individuals who falsely report (not) receiving training in the past are more likely to falsely report (not) receiving training today, then e_t and ω_t will be positively correlated. As seen from equation (11), this positive correlation will lead to a downward bias in the instrumental variable estimates of the training coefficients in column 4 of table 4.

We can eliminate the downward bias due to serial correlation in the misreporting of training over time simply by dropping past training from the first stage training incidence equation. Column 5 of table 4 reports the estimated wage growth equation when past training is omitted from the first stage training incidence equation. Comparing columns 4 and 5, we see that, as expected, dropping past training from the first stage training incidence equation leads to a moderate increase in all of the estimated training coefficients in the second stage wage growth equation. Henceforth, our discussion will refer to the estimates in column 5, so that we may assume that $\mathrm{cov}(e_{it},\omega_{it}) = 0$.

Unfortunately, the second source of bias in the instrumental variable training coefficient is not so readily eliminated. This bias stems from a non-zero correlation between the true value of training T_{it}^* and the measurement error ω_{it}. In the case of classical measurement error, $\mathrm{cov}(T_{it}^*,\omega_{it})=0$ and the instrumental variable coefficient is unbiased. However, with a discrete variable such as training incidence, the assumption that T_{it}^* and ω_{it} are uncorrelated cannot hold.[21] To see this, note that if an individual receives training so that $T_{it}^*=1$, then ω_{it} can only take on the values {0,-1}. And if an individual does not receives training so that $T_{it}^*=0$, then ω_{it} can only take on the values {0,1}. It follows immediately that T_t^* and ω_t must be *negatively correlated*. Thus, while measurement error leads to a downward bias in the OLS training coefficients, it leads to an *upward bias* in the IV training coefficients. To show this formally, we need merely simplify equation (11) as:

$$\text{(12)} \quad \text{plim } \hat{\beta}_t^{IV} = \left(1 - \left[\frac{\text{cov}(T_{it}^*, \omega_{it})}{\text{var}(T_{it}^*) + \text{cov}(T_{it}^* \omega_{it})}\right]\right)\beta_t$$

$$= \left(1 - \left[\frac{\Pr(T_{it}^* = 1, T_{it} = 0)) + \Pr(T_{it}^* = 1)\{\Pr(T_{it}^* = 0, T_{it} = 1\} - \Pr(T_{it}^* = 1, T_{it} = 0)}{\Pr((T_{it}^* = 1)\Pr(T_{it} = 1) - \Pr(, T_{it} = 1)}\right]\right)\beta_t$$

$$= \left(\frac{1}{1 - \alpha_{0t} - \alpha_{1t}}\right)\beta_t .$$

Assuming that $\alpha_{0t} + \alpha_{1t} < 1$, it follows immediately from (12) that the instrumental variable estimates of the training coefficients are upward biased in the presence of measurement error.[22]

From equation (12), we see that the proportional upward bias in the instrumental variable estimate is $1/(1-\alpha_{0t}-\alpha_{1t})$. If we had consistent estimates for α_{0t} and α_{1t}, say $\hat{\alpha}_{0t}$ and $\hat{\alpha}_{1t}$, we could obtain a consistent estimate of β_t simply by multiplying the instrumental variable estimate by $(1-\hat{\alpha}_{0t} - \hat{\alpha}_{1t})$. But if we do not observe true training, how can we estimate α_{0t} and α_{1t}?

One possible approach would be to use external information on misclassification rates, such as might be provided by a validation study. Indeed, Card (1996) takes this approach in his recent analysis of the wage effects of unions.[23] An alternative approach is to note that the greater are the

---

[21] We are indebted to Harley Frazis for initially pointing this out and for helping us to simplify the derivation of the IV bias.

[22] If $\alpha_{0t} + \alpha_{1t} > 1$, then the measurement error in reported training is so severe that either a) someone who reports receiving training is more likely not to have received training than to have received training or b) someone who reports not having received training is more likely to have received training than not to have received training. Not surprisingly, in this case, the estimated IV training coefficient will not even have the right sign.

[23] Interestingly, Card finds that "union status misclassification errors lead to a 50-75 percent attenuation in the average wage changes of observed union joiners and leavers, relative to the true wage changes of actual joiners and leavers."

mismeasurement probabilities α_{0t} and α_{1t}, the more severe is the downward bias in the OLS training coefficient and the upward bias in the instrumental variable coefficient. Thus, the ratio of the OLS and the IV training coefficients contains information about the size of α_{0t} and α_{1t}. Note that since ν_t and η_t can be expressed in terms of p_t, α_{0t} and α_{1t}, both the OLS coefficient in equation (8) and the IV coefficient in equation (12) are functions of β_t, p_t, α_{0t} and α_{1t}, and the proportional belated information bias b_t. If one assumes that measurement error has an expected value of 0, then one can solve for α_{0t} and α_{1t} as functions of the training probability p_t, the proportional belated information bias b_t, and the expected values of the training coefficients $\hat{\beta}_t^{OLS}$ and $\hat{\beta}_t^{IV}$. After the appropriate calculations, the measurement error parameters $\{\alpha_{0t}, \alpha_{1t}\}$ can be written as:

$$\text{(13)} \qquad \{\alpha_{0t}, \alpha_{1t}\} = \left\{ p_t \left[1 + \frac{b_t}{2} - \sqrt{\frac{b_t^2}{4} + \frac{1}{r_t}} \right], (1-p_t) \left[1 + \frac{b_t}{2} - \sqrt{\frac{b_t^2}{4} + \frac{1}{r_t}} \right] \right\},$$

where $r_t \equiv \text{plim} \dfrac{\hat{\beta}_t^{IV}}{\hat{\beta}_t^{OLS}}$.[24]

---

24 If we do not assume something about the expected value of measurement error or the value of either α_{0t} or α_{1t}, we do not have enough information to solve for the key parameters of interest. Specifically, our model reduces to the following two equations: $E(\omega) = [\alpha_0(1\text{-}p)\text{-}\alpha_1 p]/(1\text{-}\alpha_0\text{-}\alpha_1)$ and $(1-\alpha_0-\alpha_1)\left[1 - \frac{\alpha_1}{1-p}\left(\frac{p-\alpha_0}{1-\alpha_0-\alpha_1}\right) - \frac{\alpha_0}{p}\left(1 - \frac{p-\alpha_0}{1-\alpha_0-\alpha_1}\right) + b\right] = 1/r$. Assuming that $E(\omega)=0$ allows us to solve for α_0 and α_1 as functions of r and b. In a similar vein, Card (1996) closes his measurement error model by assuming that $\alpha_0=\alpha_1$ (in Card's notation, $q_0=1\text{-}q_1$). Our empirical results for the returns to training turn out not to be very sensitive to the assumption that the expected value of measurement error is zero. Assuming that the two misclassification parameters are identical yields similar estimates of β_t, as do other plausible assumptions. We choose to report estimates based on the assumption that $E(\omega) = 0$ since this is equivalent to assuming that $p_t=p_t^*$, whereas the alternative assumption that $\alpha_{0t}=\alpha_{1t}$ results in p_t^* estimates of .08 and .10 in the second

Substituting (13) into (12) yields:

$$(14) \quad \operatorname{plim} \hat{\beta}_t^{IV} = \left(\frac{1}{\sqrt{\frac{b_t^2}{4} + \frac{1}{r_t}} - \frac{b_t}{2}} \right) \beta_t .$$

Let $\hat{r}_t = \frac{\hat{\beta}_t^{IV}}{\hat{\beta}_t^{OLS}}$ denote the ratio of the estimated instrumental variable and OLS training coefficients

and let

$$(15) \quad \hat{\beta}_t \equiv \left(\sqrt{\frac{b_t^2}{4} + \frac{1}{\hat{r}_t}} - \frac{b_t}{2} \right) \hat{\beta}_t^{IV} \ .$$

Conditional on the belated information bias b_t, $\hat{\beta}_t$ is clearly a consistent estimator of β_t. It is straightforward to show that $\hat{\beta}_t^{IV}$ is asymptotically normally distributed with variance

$$(16) \quad V_t = \frac{1}{n_t} \frac{\operatorname{Var}\{(Z_{it} - \overline{Z}_t)\varphi_{it}\}}{\left(p\lim \frac{\sum_{i=1}^{n_t} (Z_{it} - \overline{Z}_t)(T_{it} - \overline{T}_t)}{n_t} \right)^2}$$

$$= \frac{1}{n_t} \frac{\operatorname{Var}\{(Z_{it} - \overline{Z}_t)\varphi_{it}\}}{(\operatorname{cov}(Z_{it}, T_{it}))^2},$$

and third years of tenure. Since these latter estimates are roughly half the reported probability of training, we regard them as implausibly low.

where n_t is the number of individuals in the sample with t years of tenure, $\overline{Z}_t \equiv \sum_{i=1}^{n_t} Z_{it}$, and

$\overline{T}_t \equiv \sum_{i=1}^{n_t} T_{it}$.[25] Thus, $\hat{\beta}_t$ is asymptotically normal distributed with mean β_t and variance

$$V_t\left(\sqrt{\frac{b_t^2}{4}+\frac{1}{\hat{r}_t}}-\frac{b_t}{2}\right)^2 .$$

For expositional convenience, we have thus far assumed that training is the only explanatory variable in the wage growth equation. It is straightforward to extend the analysis so as to allow additional explanatory variables. Specifically, suppose that instead of (6'), the wage growth equation is the more general equation (6). Adopting a similar argument to that in Card (1996), one can then show that

$$\text{(8')} \qquad \operatorname{plim}\hat{\beta}_t^{OLS} = \frac{[1-(\nu_t+\eta_t)]\beta_t - \dfrac{R_t^2}{1-\alpha_1-\alpha_0}}{R_t^2} + b_t ,$$

where R_t^2 denotes the theoretical R-squared coefficient from a linear regression of T_{it} against the other explanatory variables in the wage growth equation, and the belated information bias is now given by $b_t = \left[\frac{\operatorname{cov}(T_{it},\theta_{it}|\Delta X_{it})}{\operatorname{var}(T_{it}|\Delta X_{it})}\right]\left(\frac{1}{\beta_t}\right)$.[26] The appendix demonstrates that the insrumental variable

---

[25] Note that in the standard instrumental variable formulation, we would have $V_t = \frac{1}{n_t}\frac{\operatorname{Var}\{(Z_{it}-\overline{Z}_t)\varphi_{it}\}}{(\operatorname{cov}(Z_{it},T_{it}))^2} = \frac{1}{n_t}\frac{[\operatorname{Var}(Z_{it}-\overline{Z}_t)]\operatorname{var}(\varphi_{it})\}}{(\operatorname{cov}(Z_{it},T_{it}))^2}$. In our current formulation, we cannot carry out the last step because Z_{it} and φ_{it} are not independent.

[26] The derivation of (8') assumes T_{it} is the only explanatory variable in the wage growth equation that is measured with error. The derivation also assumes that $\operatorname{cov}(\Delta X_{it},\omega_{it}| T_{it}^*) = 0$. Our derivation differs from Aigner's in that Aigner assumes that $\operatorname{cov}(\Delta X_{it}, \omega_{it}) = 0$ even though ΔX_{it} may be correlated with T_{it} and T_{it}^*.

bias does not change when additional explanatory variables are included in the equation. Using (8') and (12) to solve for α_{0t} and α_{1t}, one obtains

$$(13') \qquad \alpha_{0t} = p_t \left[1 + \frac{b_t}{2}(1 - R_t^2) - \sqrt{\frac{b_t^2}{4}(1 - R_t^2)^2 + \frac{1}{r_t}(1 - R_t^2) + R_t^2} \right],$$

$$\alpha_{1t} = (1 - p_t) \left[1 + \frac{b_t}{2}(1 - R_t^2) - \sqrt{\frac{b_t^2}{4}(1 - R_t^2)^2 + \frac{1}{r_t}(1 - R_t^2) + R_t^2} \right].$$

Thus, letting $\hat{R}_t^2$ denote the estimated R-squared coefficient from a linear regression of observed training against the other explanatory variables in the wage growth equation,

$$(15') \; \hat{\beta}_t \equiv \left(\sqrt{\frac{b_t^2}{4}(1 - 2\hat{R}_t^2 + (\hat{R}_t^2)^2) + \frac{1}{\hat{r}_t}(1 - \hat{R}_t^2) + \hat{R}_t^2} - \frac{b_t}{2}(1 - \hat{R}_t^2) \right) \hat{\beta}_t^{IV}$$

is a consistent estimator of β_t.

The estimated returns to training that result from the calculations described above are presented in Table 5.[27] In interpreting this table, recall that these estimates depend on the magnitude of the belated information bias, b_t. Column 1 indicates the parameter estimates when there is no belated information. Columns 2 and 3 indicate the parameter estimates that result when b_t is .05 and .1, respectively.

---

[27] As noted if footnote 26, the derivations of (8') and (15') assume that T_{it} is the only variable in the wage growth equation that is measured with error. The estimates in Table 5 are therefore obtained from single year OLS and IV equations instead of the pooled equations presented in Table 4. The pooled and single year estimates are very close.

When there is no belated information bias, the estimated returns to training in column 1 are .038 in the second year of tenure, .093 in the third year of tenure, and .069 in the fourth year of tenure. An increase in the belated information bias results in a relatively small drop in the estimated returns to training. For example, an increase in b_t from 0 to 10 percent causes the estimated returns for second and third year training to fall by about 10 percent and the estimated return for fourth year training to fall by about 6 percent.

The corrected wage coefficients in Table 5 are considerably lower than the IV coefficients in column 5 of table 4. Nevertheless, the corrected coefficients still indicate that the effect of training on wages is quite substantial. Most notably, training in the third or fourth year of tenure increases wage growth during that year by somewhere between 6 and 9 percent. The effect of training becomes even more striking when one considers our finding in the previous section that training incidence is correlated over time for a given individual. If an individual receiving training in the third year also received training in the second year of tenure, his total wage growth from year 1 to year 3 is 12 to 13 percent higher than an individual who did not receive training in either year.

Besides presenting the estimated returns to training, we also present our estimates of the measurement error parameters α_{0t} and α_{1t} in table 5.[28] Note that these estimates, like the estimated returns to training, depend on the magnitude of the belated information bias. When there is no belated information bias and a full set of explanatory variables are included in the wage growth equation, $\alpha_{0t} \equiv \Pr(T_{it}=1|T_{it}^*=0)$ is estimated to be .08 for t=2, .08 for t=3, and .02 for t=4. Similarly, $\alpha_{1t} \equiv \Pr(T_{it}=0|T_{it}^*=1)$ is estimated to be .377 for t=2, .358 for t=3, and .09 for t=4. In other words, during the second and third years of tenure, eight percent of persons who did not

---

[28] Let $\hat{p}_t$ denote the proportion of individuals in the sample with tenure t who report receiving training. Substituting $\hat{p}_t$ and $\hat{r}_t = \frac{\hat{\beta}_t^{IV}}{\hat{\beta}_t^{OLS}}$ for p_t and r_t in equation (13') yields consistent estimates $\hat{\alpha}_{0t}$ and $\hat{\alpha}_{1t}$ of the measurement error parameters α_{0t} and α_{1t}.

receive training report that they did receive training, and roughly a third of those who received training reported that they did not receive training. These proportions of misreporting fall sharply in the fourth year of tenure.

Our estimates thus indicate that there is a sizable amount of misreporting in the NLSY training data. One may ask whether these measurement error estimates are reasonable. Barron, Berger, and Black (1994b) is the only study we know of that has sought to quantify the measurement error in training. Although their survey is different from the NLSY data used here, it is still instructive to compare their results with ours.[29] In the Barron-Berger-Black data, employers and employees report similar amounts of training: thirty percent of workers and thirty-three percent of employers report formal training. However, while the means are similar, the correlation between worker reported training and employer reported training is only .369. If the employer's response were always correct but the worker sometimes misreported, then it can be shown that the Barron-Berger-Black data imply that α_0 = Pr(Worker reports training|Employer does not report training) = .183 and α_1 = Pr(Worker does not report training|Employer reports training) = .457. These measurement error estimates are even larger than ours, although it should be kept in mind that the reported probability of training is almost twice as high in the Barron-Berger-Black data as in the NLSY data. Alternatively, if one assumes that workers and employers are equally likely to misreport training, then it can be shown that the Barron-Berger-Black data imply that $\alpha_0 = .125$ and $\alpha_1 = .27$. These estimates are similar to those obtained here; at the very least, our estimates do not look unreasonable.

---

[29] Barron, Berger, and Black ask both an employee and his employer about on-site formal training, off-site formal training, and various types of informal training during the first two weeks of employment. See Barron, Berger, and Black for a more complete description and see Loewenstein and Spletzer (1994) for a further comparison of "EOPP-like" surveys and the NLSY. We are grateful to Dan Black for providing us the figures for aggregate "formal training."

Finally, any conclusions with respect to the human capital model require unbiased tenure coefficients. Of course, a bias in the estimated training coefficients will also lead to a bias in all of the other variables in the wage growth regression, including the tenure-specific intercepts. Having obtained a consistent estimate, $\hat{\beta}_t$, of the training coefficient, we can obtain consistent estimates of the coefficients on the remaining variables by estimating the equation

$$(17) \quad \Delta w_t - \hat{\beta}_t T_{it} = (\gamma_t - \gamma_{t-1}) + c_3 \Delta X_{it} + \psi_t,$$

where $\psi_t \equiv \theta_{it} + \Delta\varepsilon_{it} - \beta_t \omega_{it} + (\beta_t - \hat{\beta}_t) T_{it}$. The resulting tenure coefficients are presented in the lower panel of table 5. As is evident from a comparison of tables 4 and 5, the consistently estimated tenure coefficients are not substantially different from the instrumental variable tenure coefficients.

Using these unbiased estimates of the training coefficients and the tenure coefficients, the fact that the training coefficients are substantially larger than the corresponding tenure coefficients in the later years of tenure provide strong support for the human capital model. Between the first and second year of tenure, the wages of workers who do not receive training grow by approximately three and one-half percent, but the wages of workers who receive training grow by approximately seven percent. Individuals who do not receive training between their second and third years of tenure experience no wage growth, whereas workers who receive training in their third year of tenure experience wage growth between eight and nine percent. Similarly, individuals who do not receive training in their fourth year of tenure receive essentially no wage growth, whereas workers who receive training in their fourth year of tenure experience wage growth of approximately six to seven percent. Interestingly, unexplained wage growth is greatest between the first and second year of tenure. We suspect that this may reflect the fact that the NLSY training questions likely miss most informal training, training which is likely to be most important early in the employment relationship.[30]

VI. Conclusions and Discussion

The human capital model attributes rising wage-tenure and wage-experience profiles to investments in worker productivity. Therefore, an implicit assumption in the human capital model is that training must be occurring throughout a worker's stay at an employer. There are at least two factors that may lead to significant amounts of belated training. First, workers are often better able to absorb expensive training after an initial period of "learning by doing" in which they have been acclimatized to their job and work environment. Second, information about the quality of a firm-worker match is often revealed only after a period of time. An employer who delays training can lower the probability of investing in bad matches.

The data presented in section II indicate quite clearly that a substantial amount of training occurs after the first year of employment. Annual training incidence nearly doubles from 9.89 percent in the first year of tenure to 17.21 percent in the second year of tenure, and is then roughly constant around 18 percent in the third and fourth years of tenure. The data also indicate that there are significant match specific fixed effects in training: individuals who have received previous training at their current employer are more likely to receive current training.

Section III of the paper examined the relationship between training and wages. Taking first differences for job stayers eliminates potential biases stemming from unobserved fixed individual and match effects, but does not eliminate the bias from belated information. Specifically, if an

---

[30] Since there is a positive correlation between the receipt of formal training and the receipt of informal training (see Loewenstein and Spletzer (1994) for evidence on this score), the coefficient on formal training should partly pick up the returns to informal training. If informal training is more important early in the employment relationship, then this by itself should cause the coefficient on second year training to exceed those on third and fourth year training. Our results indicate that this effect is more than offset by higher returns to belated training.

employer belatedly offers a worker a raise and invests in his training upon discovering that his match-specific productivity is high, then the training coefficient in the fixed effects wage equation will still be biased upward. To eliminate the bias from belated information, we have estimated the fixed effects equation with a two stage procedure that uses previous training and fixed individual and match-specific variables as instruments.

The instrumental variable technique is also the standard method of correcting for measurement error. While the potential upward bias in the OLS training coefficient caused by a likely positive correlation between unobserved worker ability and training has been the object of a fair amount of discussion, the downward bias in the OLS training coefficient caused by measurement error has received little, if any, attention.[31] Recent research suggests that measurement error in education may cause the estimated returns to education to be biased downward by as much as 15%. One suspects that, if anything, training is more difficult to measure than education.

Indeed, our instrumental variables estimation yields substantially higher estimated returns to training than does our OLS estimation. However, instrumental variables estimation does not yield consistent estimates when the explanatory variable measured with error is dichotomous, or more generally, when the explanatory variable measured with error can take on only a limited range of values. The difficulty stems from the fact that the measurement error in such a variable cannot be classical, but must be negatively correlated with the variable's true value. In our application, the fact that the instrumented training incidence measure is negatively correlated with the measurement

---

[31] A notable exception here is Brown (1989), who uses training data from the Panel Study of Income Dynamics. This survey asks a worker how long it would take the average person to become fully trained and qualified in a job like his, and Brown assumes an individual to be receiving training if his job tenure is less than the time it takes to become fully trained and qualified. Brown points out that the length of a training spell may be measured with error because an individual's initial set of skills may differ from that of the "average person" and because it is not clear whether a respondent interprets "job" as referring to the current position [as Brown assumes], the current firm, or possibly the current occupation. Interestingly, Brown (1994) demonstrates that measurement error may actually be inversely correlated with observed training, causing the estimated returns to training to be biased upward.

error leads to an upward biased estimate of the returns to training. The relative magnitudes of the IV and OLS training coefficients provide information about the size of measurement error bias. Our results indicate that this bias is substantial: if there is no belated information bias, the IV training coefficients are biased upward by as much as 75%.

After correcting for measurement error bias, our estimated wage growth equation provides strong support for the human capital model. The training coefficients are positive and are substantially larger than the corresponding tenure coefficients, particularly, for the third and fourth years of tenure. Between the first and second year of tenure, the wages of workers who do not receive training grow by approximately three and one-half percent, but the wages of workers who receive training grow by approximately seven percent. And while individuals who do not receive training experience basically no wage growth between the second and fourth years of tenure, workers who receive training in either the third or fourth years of tenure experience wage growth of approximately six to nine percent.

Finally, we may note that it is quite possible that the unexplained wage growth between the first and second years of tenure is itself due to human capital accumulation. The training question in the 1988-1991 NLSY surveys appears to primarily measure formal training. The NLSY did not begin explicitly asking about informal training until 1993. Loewenstein and Spletzer (1994) provide evidence that the training measure in the earlier surveys misses most informal training, Thus, our present finding that unexplained wage growth is greatest between the first and second year of tenure may simply reflect the fact that informal training is most important early in the employment relationship. A thorough investigation of this hypothesis must be deferred until future NLSY surveys using the new informal training questions become available.

References

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

Aigner, Dennis J. 1973. "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics*, pp. 49-60.

Altonji, Joseph G. and Robert A. Shakotko. 1987. "Do Wages Rise with Job Seniority?" *Review of Economic Studies*, pp. 437-459.

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*, pp. 58-79.

Ashenfelter, Orley and Alan Krueger. 1994. "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review*, pp. 1157-1173.

Barron, John M., Mark C. Berger, and Dan A. Black. 1994a. "Do Workers pay for On-the-Job Training?" Unpublished paper, University of Kentucky.

Barron, John M., Mark C. Berger, and Dan A. Black. 1994b. "How Well do We Measure Training?" Unpublished paper, University of Kentucky.

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

Barron, John M., Dan A. Black, and Mark A. Loewenstein. 1993. "Gender Differences in Training, Capital, and Wages." *Journal of Human Resources*, pp. 343-364.

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

Bartel, Ann P. and Nachum Sicherman. 1995. “Technological Change and the Skill Acquisition of Young Workers.” NBER Working Paper # 5107.

Becker, Gary S. 1962. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy*, pp. 9-49.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, pp. 352-367.

Blinder, Alan S. and Yoram Weiss. 1976. "Human Capital and Labor Supply: A Synthesis." *Journal of Political Economy*, pp. 449-472.

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*, pp. 971-991.

Brown, James N. 1994. "Classification Error in Linear Spline Regressions." Unpublished paper, Rice University.

Card, David. 1995. "Earnings, Schooling, and Ability Revisited." *Research in Labor Economics*, pp. 23-48.

Card, David. 1996. “The Effect of Unions on the Structure of Wages: A Longitudinal Analysis.” *Econometrica*, pp. 957-979.

Duncan, Greg J. and Saul Hoffman. 1979. "On-the-Job Training and Earnings Differences by Race and Sex." *The Review of Economics and Statistics*, pp. 594-603.

Freeman, Richard B. 1984. "Longitudinal Analyses of the Effects of Trade Unions." *Journal of Labor Economics*, pp. 1-26.

Johnson, William R. 1978. "A Theory of Job Shopping." *Quarterly Journal of Economics*, pp. 261-278.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy,* pp. 972-990.

Heckman, James J. 1978 “Dummy Endogenous Variables in a Simultaneous Equation System.” *Econometrica*, pp. 931-958.

Kelejian, Harry H. 1971. "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables," *Journal of the American Statistical Association*, pp. 373-374.

Lazear, Edward P. 1981. "Agency, Earnings Profiles, Productivity, and Hours Restrictions," *American Economic Review*, pp. 606-620.

Lillard, Lee A. and Hong W. Tan. 1992. Private Sector Training: Who Gets It and What Are Its Effects? *Research in Labor Economics*, pp. 1-62.

Loewenstein, Mark A. and James R. Spletzer. 1994. "Informal Training: A Review of Existing Data and Some New Evidence." Bureau of Labor Statistics Working Paper 254.

Loewenstein, Mark A. and James R. Spletzer. 1996a. "Delayed Formal On-the-Job Training." Unpublished Working Paper, Bureau of Labor Statistics.

Loewenstein, Mark A. and James R. Spletzer. 1996b. "Dividing the Costs and Returns to General Training." Unpublished Working Paper, Bureau of Labor Statistics.

Lynch, Lisa M. 1991. "The Role of Off-the-Job vs. On-the-Job Training for the Mobility of Women Workers." *American Economic Review*, pp. 151-156.

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

Mincer, Jacob. 1962. "On-the-Job Training: Costs, Returns, and Some Implications." *Journal of Political Economy*, pp. 50-79.

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

Parsons, Donald O. 1989. "On-the-Job Learning and Wage Growth." Unpublished paper, Ohio State University.

Pergamit, Michael R. and Janice Shack-Marquez. 1987. "Earnings and Different Types of Training." Bureau of Labor Statistics Working Paper #165.

Topel, Robert. 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority." *Journal of Political Economy*, pp. 145-176.

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

Table 1: Training-Tenure Relationship, 1988-1991 NLSY

Tenure	Sample Size	Percent Trained
1	11014	9.89%
2	3179	17.21%
3	1162	18.16%
4	388	18.30%

Table 2: Training Incidence

Probability	Cell Size	(1)	(2)	(3)
$Pr(T_1=1\|PCJ_1=0)$	11014	.0989 * (.0028)		
$Pr(T_1=1\|PCJ_1=0,POJ_1=0)$	2763		.0894 * (.0054)	.0689 (.0406)
$Pr(T_1=1\|PCJ_1=0,POJ_1>0)$	386		.2876 * (.0230)	.2362 * (.0433)
$Pr(T_2=1\|PCJ_2=0)$	2753	.1344 * (.0065)		
$Pr(T_2=1\|PCJ_2=0,POJ_2=0)$	655		.1267 * (.0130)	.0941 * (.0419)
$Pr(T_2=1\|PCJ_2=0,POJ_2>0)$	62		.0968 * (.0376)	.0589 (.0566)
$Pr(T_2=1\|PCJ_2=1)$	343	.3061 * (.0249)		
$Pr(T_2=1\|PCJ_2=1,POJ_2=0)$	80		.2500 * (.0484)	.1757 * (.0532)
$Pr(T_2=1\|PCJ_2=1,POJ_2>0)$	30		.3333 * (.0861)	.2732 * (.0706)
$Pr(T_3=1\|PCJ_3=0)$	850	.1176 * (.0110)		
$Pr(T_3=1\|PCJ_3=0,POJ_3=0)$	159		.0881 * (.0225)	.0625 (.0474)
$Pr(T_3=1\|PCJ_3=0,POJ_3>0)$	6		.0000 (.0000)	-.0523 (.1333)
$Pr(T_3=1\|PCJ_3=1)$	228	.2807 * (.0298)		
$Pr(T_3=1\|PCJ_3=1,POJ_3=0)$	42		.2381 * (.0657)	.1713 * (.0630)
$Pr(T_3=1\|PCJ_3=1,POJ_3>0)$	5		.2000 (.1789)	.1901 (.1603)
$Pr(T_3=1\|PCJ_3=2)$	51	.4510 * (.0697)		
$Pr(T_3=1\|PCJ_3=2,POJ_3=0)$	6		.3333 * (.1924)	.2353 (.1338)
$Pr(T_3=1\|PCJ_3=2,POJ_3>0)$	2		1.0000 * (.0000)	.9294 * (.2235)
$Pr(T_4=1\|PCJ_4=0)$	251	.1116 * (.0199)	.1116 * (.0199)	.0761 (.0452)
$Pr(T_4=1\|PCJ_4=1)$	84	.2143 * (.0448)	.2143 * (.0448)	.1288 * (.0526)
$Pr(T_4=1\|PCJ_4=2)$	30	.4000 * (.0894)	.4000 * (.0894)	.2979 * (.0699)
$Pr(T_4=1\|PCJ_4=3)$	10	.3000 * (.1449)	.3000 * (.1449)	.2154 * (.1069)
Explanatory Variables		No	No	Yes

1988-1991 NLSY. Sample size = 15743.
Columns 1-2: Maximum likelihood estimates. Column 3: OLS coefficients. Standard errors in parentheses. * implies statistically different from zero at the 5% level of significance (one tailed test).
The statistics reported above are only for those individuals who did not have an ongoing training spell (PCJ>0) at the time of the previous interview. Statistics are not reported for individuals who had a training spell ongoing at the date of the last interview. Statistics are not reported in columns 2-3 for individuals

who did not have a Prior Other Job observed.

Explanatory variables are year dummies, AFQT, race, gender, age, marital status, urban residence, local area unemployment rate, SMSA, union, experience and experience squared, firm size, education, school attendance in the previous year, multiple site firm, number of previous jobs, government employment, part-time employment, number of children, and industry and occupation.

Table 3: Wage Level Regressions (Job Stayers)

	Mean	(1)	(2)	(3)
Tenure=2	.6701	1.9936 *	1.9223 *	1.7761 *
		(.0085)	(.0158)	(.0883)
Tenure=3	.2477	2.0503 *	1.9378 *	1.7449 *
		(.0140)	(.0205)	(.0906)
Tenure=4	.0821	2.1313 *	1.9730 *	1.7443 *
		(.0243)	(.0328)	(.0990)
Tenure=1 x Training	.1329		.1607 *	.0554 *
			(.0209)	(.0169)
Tenure=2 x Training	.1789		.1246 *	.0628 *
			(.0196)	(.0158)
Tenure=3 x Training	.0630		.1993 *	.0773 *
			(.0325)	(.0259)
Tenure=4 x Training	.0151		.1535 *	.0739
			(.0672)	(.0533)
# Training Spells, Previous Jobs	.0302		.1203 *	.0576
			(.0383)	(.0307)
Explanatory Vars.		No	No	Yes

1988-1991 NLSY. OLS Regression Coefficients. Standard errors in parentheses. Dependent variable is the log real wage. Mean (Standard deviation) of dependent variable is 2.0190 (.4659). Sample Size = 4432 (Job Stayers, Tenure$\geq$2).

* implies statistically different from zero at the 5% level of significance (two tailed test).

Equations 2-3 include controls for whether the training in the current year is ongoing at the date of the interview, and controls for whether no training is observed in previous jobs because of left censored data.

Explanatory variables are year dummies, AFQT, race, gender, age, marital status, urban residence, local area unemployment rate, SMSA, union, experience and experience squared, firm size, education, school attendance in the previous year, multiple site firm, number of previous jobs, government employment, part-time employment, number of children, and industry and occupation.

Table 4: Within Job Wage Growth Regressions

	Mean	(1)	(2)	(3)	(4)	(5)
Tenure=2	.6701	.0491 *	.0458 *	.0379 *	.0297	.0277
		(.0052)	(.0057)	(.0127)	(.0156)	(.0164)
Tenure=3	.2477	.0259 *	.0187 *	.0068	-.0084	-.0128
		(.0085)	(.0094)	(.0144)	(.0172)	(.0187)
Tenure=4	.0821	.0204	.0090	.0110	.0054	.0029
		(.0147)	(.0163)	(.0191)	(.0229)	(.0243)
Tenure=2 x Training	.1171		.0171	.0190	.0630	.0666
			(.0152)	(.0152)	(.0515)	(.0537)
Tenure=3 x Training	.0454		.0520 *	.0482 *	.1285 *	.1412 *
			(.0240)	(.0239)	(.0610)	(.0692)
Tenure=4 x Training	.0151		.0575	.0600	.0814	.0837
			(.0418)	(.0417)	(.0851)	(.0894)
Predicted (OLS)						
Tenure x Training		No	No	No	Yes	Yes
1st Stage Variables					$T_{t-\tau}$,X	X
Explanatory Vars.		No	No	Yes	Yes	Yes

1988-1991 NLSY. Dependent variable is log real wage growth. Mean (Standard deviation) of dependent variable is 0.0410 (.2809). Sample Size = 4432. OLS Regression Coefficients. Columns 1-3: OLS Standard errors in parentheses. Columns 4-5: 2SLS Standard errors in parentheses. * implies statistically different from zero at the 5% level of significance (two tailed test).

Equations 2-5 include controls for whether the training in the current year is ongoing at the date of the interview and whether training was ongoing at the date of the previous interview.

1st stage variables "$T_{t-\tau}$" are indicators of training in previous years of tenure, number of previous training spells in other jobs, controls for whether the training in the previous year was ongoing at the date of the interview, and controls for whether no training is observed in previous jobs because of left censored data. 1st stage variables "X" are an intercept, AFQT, race, gender, age, marital status, urban residence, local area unemployment rate, SMSA, union, experience and experience squared, firm size, education, school attendance in the previous year, multiple site firm, number of previous jobs, government employment, part-time employment, number of children, and industry and occupation.

Explanatory variables are (in first differences) year dummies, marital status, local area unemployment rate, experience squared, education, part-time employment, school attendance in the previous year, and number of children.

Table 5: Derivation of Unbiased Coefficients

	(1)	(2)	(3)
b_2 Belated Info.Bias, t=2	0	.05	.1
b_3 Belated Info.Bias, t=3	0	.05	.1
b_4 Belated Info.Bias, t=4	0	.05	.1
$\{\alpha_{02},\alpha_{12}\}$	{.080,.377}	{.084,.397}	{.088,.416}
$\{\alpha_{03},\alpha_{13}\}$	{.080,.358}	{.084,.378}	{.089,.397}
$\{\alpha_{04},\alpha_{14}\}$	{.020,.090}	{.025,.11}	{.029,.129}
Tenure=2 x Training (β_2)	0.038 (0.032)	0.036 (0.031)	0.034 (0.030)
Tenure=3 x Training (β_3)	0.093 * (0.031)	0.089 * (0.030)	0.084 * (0.028)
Tenure=4 x Training (β_4)	0.069 (0.050)	0.067 (0.049)	0.065 (0.047)
Tenure=2 ($\gamma_2 - \gamma_1$)	0.033 * (0.012)	0.034 * (0.012)	0.034 * (0.012)
Tenure=3 ($\gamma_3 - \gamma_2$)	-0.004 (0.014)	-0.003 (0.014)	-0.002 (0.014)
Tenure=4 ($\gamma_4 - \gamma_3$)	0.006 (0.018)	0.007 (0.018)	0.007 (0.018)

* implies statistically different from zero at the 5% level of significance (two tailed test).

Appendix for "Belated Training: The Relationship Between Training, Tenure, and Wages"

This appendix shows how to generalize the analysis of measurement error in Section 4 so as to include additional explanatory variables in the wage growth equation. To make the analysis tractable, we assume that only one variable is measured with measurement error. Thus, instead of considering an equation that pools all of the observations, we consider separate equations for each year of tenure. These are the equations that are used to obtain the coefficients reported in Table 5.

When training is the only explanatory variable in the wage growth equation,

$$\text{(A1)} \qquad \operatorname{plim} \hat{\beta}_t^{OLS} = \beta_t \left(1 - \frac{\operatorname{cov}(T_{it}, \omega_{it})}{\operatorname{var}(T_{it})} \right) + \frac{\operatorname{cov}(T_{it}, \theta_{it})}{\operatorname{var}(T_{it})}$$

$$= \beta_t \left(\frac{\operatorname{cov}(T_{it}^*, T_{it})}{\operatorname{var}(T_{it})} \right) + \frac{\operatorname{cov}(T_{it}, \theta_{it})}{\operatorname{var}(T_{it})} .$$

When the vector ΔX_t of additional variables is included in the wage growth equation along with training, the expected value of the OLS training coefficient is given by

$$\text{(A2)} \qquad \operatorname{p\,lim} \hat{\beta}_t^{OLS} = \beta_t \left(\frac{\operatorname{cov}(T_{it}^*, T_{it} | \Delta X_{it})}{\operatorname{var}(T_{it} | \Delta X_{it})} \right) + \frac{\operatorname{cov}(T_{it}, \theta_{it} | \Delta X_{it})}{\operatorname{var}(T_{it} | \Delta X_{it})}$$

Card (1996) shows that

$$\frac{\operatorname{cov}(T_{it}^*, T_{it} | \Delta X_{it})}{\operatorname{var}(T_{it} | \Delta X_{it})} = \frac{\dfrac{\operatorname{cov}(T_{it}^*, T_{it})}{\operatorname{var}(T_{it})} - \dfrac{R_t^2}{1 - \alpha_1 - \alpha_0}}{1 - R_t^2} ,$$

where R_t^2 denotes the theoretical R-squared coefficient from a linear regression of T_{it} against he other explanatory variables in the wage growth equation. Thus, (A2) can be rewritten as

$$\text{(A3)} \qquad p\lim \hat{\beta}_t^{OLS} = \beta_t \left(\frac{\frac{\text{cov}(T_{it}*, T_{it})}{\text{var}(T_{it})} - \frac{R_t^2}{1-\alpha_1-\alpha_0}}{1-R_t^2} \right) + \frac{\text{cov}(T_{it}, \theta_{it} | \Delta X_{it})}{\text{var}(T_{it} | \Delta X_{it})}$$

$$= \beta_t \left(\frac{1-\eta_t-\nu_t-\frac{R_t^2}{1-\alpha_1-\alpha_0}}{1-R_t^2} + b_t \right).$$

When training is the only explanatory variable in the wage growth equation, we have

$$\text{(A4)} \qquad p\lim \hat{\beta}_t^{IV} = \beta_t \left(1 - \frac{\text{cov}(T_{it}*, \omega_{it})}{\text{var}(T_{it}*) + \text{cov}(T_{it}*, \omega_{it}))} \right)$$

$$= \beta_t \left(\frac{\text{var}(T_{it}*)}{\text{var}(T_{it}*) + \text{cov}(T_{it}*, \omega_{it}))} \right)$$

$$= \beta_t \left(\frac{\text{var}(T_{it}*)}{\text{cov}(T_{it}*, T_{it})} \right),$$

where we have used the fact that $\omega_{it} = T_{it} - T_{it}*$. When the vector ΔX_t of additional variables is included in the wage growth equation along with training, the expected value of the IV training coefficient is given by

$$\text{(A5)} \qquad p\lim \hat{\beta}_t^{IV} = \beta_t \left(\frac{\text{var}(T_{it}* | \Delta X_{it})}{\text{cov}(T_{it}*, T_{it} | \Delta X_{it})} \right).$$

Let

(A6) $\quad T_{it}^* = p_t^* + (\Delta X_{it} - \overline{\Delta X}_t)c + v_{it}$

denote the linear projection of T_t^* on ΔX_{it}, where $\overline{\Delta X}_t$ denotes the mean of ΔX_{it} and $E(v_{it}) = E(v_{it}\Delta X_{it}) = 0$. The implied linear projection of T_{it} on ΔX_{it} is

(A7) $\quad T_{it}^* = p_t + (1-\alpha_1 - \alpha_0)(\Delta X_{it} - \overline{\Delta X}_t)c + \xi_{it}$.

It is straightforward to show that ξ_{it} is orthogonal to ΔX_{it}. Combining (A5), (A6), and (A7), one obtains

$$\text{(A8)} \quad p\lim \hat{\beta}_t^{IV} = \beta_t\left(\frac{\operatorname{var}(T_{it}^*) - p_t^* - c(\Delta X_{it} - \overline{\Delta X}_t)}{\operatorname{cov}(T_{it}^* - p_t^* - c(\Delta X_{it} - \overline{\Delta X}_t), T_{it} - p_t - (1-\alpha_1-\alpha_0)c(\Delta X_{it} - \overline{\Delta X}_t))}\right)$$

$$\beta_t\left(\frac{\operatorname{var}(T_{it}^*) - c'V_{\Delta X_t \Delta X_t}c}{\operatorname{cov}(T_{it}^*, T_{it}) - (1-\alpha_1-\alpha_0)c'V_{\Delta X_t \Delta X_t}c}\right),$$

where $V_{\Delta X_t \Delta X_t}$ is the population variance-covariance matrix of ΔX_{it}. When $E(\omega_{it}) = 0$, it is straightforward to show that $\operatorname{var}(T_{it}^*) = p_t(1- p_t)$, $\operatorname{cov}(T_{it}^*, T_{it}) = p_t(1- \alpha_1 - p_t)$, and $1- \alpha_1 - \alpha_0 = (1- \alpha_1 - p_t)/(1- p_t)$. Thus, (A8) can be rewritten as

$$\text{(A9)} \quad p\lim \hat{\beta}_t^{IV} = \beta_t\left(\frac{p_t(1-p_t) - c'V_{\Delta X_t \Delta X_t}c}{p_t(1-\alpha_1-p_t) - c'V_{\Delta X_t \Delta X_t}c\left(\frac{1-\alpha_1-p_t}{1-p_t}\right)}\right)$$

$$= \beta_t\left(\frac{(1-p_t)}{1-\alpha_1-p_t}\right)$$

$$= \beta_t\left(\frac{1}{1-\alpha_1-\alpha_0}\right).$$

Now let us find the variance of $\hat{\beta}_t$. As is well known, the instrumental and two stage least squares estimators are equivalent. It is straightforward to show that

$$\text{(A20) (A20)} \quad \begin{bmatrix} \hat{\beta}_t^{TSLS} \\ \hat{c}_{31}^{TSLS} \\ \hat{c}_{32}^{TSLS} \\ . \\ \hat{c}_{3m}^{TSLS} \end{bmatrix} = \begin{bmatrix} \beta_t \\ c_{31} \\ c_{32} \\ . \\ c_{3m} \end{bmatrix} + (\mathbf{z_t}'\mathbf{z_t})^{-1}\mathbf{z_t}' \begin{bmatrix} \varphi_{1t} \\ \varphi_{2t} \\ . \\ . \\ \varphi_{n_t t} \end{bmatrix}$$

$$= \begin{bmatrix} \beta_t \\ c_{31} \\ c_{32} \\ . \\ c_{3m} \end{bmatrix} + (\frac{\mathbf{z_t}'\mathbf{z_t}}{n_t})^{-1} \begin{bmatrix} (1/n_t)\sum_{i=1}^{n_t}(Z_{it} - \overline{Z}_t)\varphi_{it} \\ (1/n_t)\sum_{i=1}^{n_t} d_{i1t}\varphi_{it} \\ . \\ . \\ (1/n_t)\sum_{i=1}^{n_t} d_{i1t}\varphi_{it} \end{bmatrix},$$

where

$$\mathbf{z}_t \equiv \begin{bmatrix} Z_{1t} & d_{11t} & d_{12t} & \dots & d_{1mt} \\ Z_{2t} & d_{21t} & d_{22t} & \dots & d_{2mt} \\ . & . & . & \dots & \\ . & . & . & \dots & \\ Z_{n_t t} & d_{n_t 1t} & d_{n_t 2t} & \dots & d_{n_t mt} \end{bmatrix}.$$

and d_{ijt} denotes the j^{th} element in the vector $\Delta X_{it} - \overline{\Delta X}_t$.

$$\sqrt{n_t}\begin{bmatrix}(1/n_t)\sum_{i=1}^{n_t}(Z_{it}-\overline{Z}_t)\varphi_{it}\\ (1/n_t)\sum_{i=1}^{n_t}d_{i1t}\varphi_{it}\\ .\\ .\\ (1/n_t)\sum_{i=1}^{n_t}d_{imt}\varphi_{it}\end{bmatrix}$$ is normal with mean $\begin{bmatrix}\operatorname{cov}(Z_{it},\varphi_{it})\\ \operatorname{cov}(d_{i1t},\varphi_{it})\\ .\\ .\\ \operatorname{cov}(d_{imt},\varphi_{it})\end{bmatrix}$ and variance covariance matrix

$$\mathbf{A}_t \equiv \begin{bmatrix}\operatorname{var}((Z_{it}-\overline{Z}_t)\varphi_{it}) & \operatorname{cov}((Z_{it}-\overline{Z}_t)\varphi_{it},d_{i1t}\varphi_{it}) & \cdots & \operatorname{cov}((Z_{it}-\overline{Z}_t)\varphi_{it},d_{imt}\varphi_{it})\\ \operatorname{cov}((Z_{it}-\overline{Z}_t)\varphi_{it},d_{i1t}\varphi_{it}) & & & \\ . & & \sigma_\varphi^2 \operatorname{plim}\frac{\mathbf{d_t}'\mathbf{d_t}}{n} & \\ . & & & \\ \operatorname{cov}((Z_{it}-\overline{Z}_t)\varphi_{it},d_{imt}\varphi_{it}) & & & \end{bmatrix}.$$

Thus, $\begin{bmatrix}\hat{\beta}_t^{TSLS}\\ \hat{c}_{31}^{TSLS}\\ \hat{c}_{32}^{TSLS}\\ .\\ \hat{c}_{3m}^{TSLS}\end{bmatrix}$ is normal with variance covariance matrix $\mathbf{V}_t = (1/n_t)\operatorname{plim}(\frac{\mathbf{z_t}'\mathbf{z_t}}{n})^{-1}\mathbf{A}_t \operatorname{plim}(\frac{\mathbf{z_t}'\mathbf{z_t}}{n})^{-1}$.

Letting $\mathbf{V}_t \equiv [v_{ijt}]$, the variance of $\hat{\beta}_t^{TSLS}$ is simply $[v_{11}^t]$. Since

$$\hat{\beta}_t \equiv \left(\sqrt{\frac{b_t^2}{4}(1-2\hat{R}^2+(\hat{R}^2)^2)+\frac{1}{\hat{r}_t}(1-\hat{R}^2)+\hat{R}^2}-\frac{b_t}{2}(1-\hat{R}^2)\right)\hat{\beta}_t^{IV}$$ and

$$\left(\sqrt{\frac{b_t^2}{4}(1-2\hat{R}^2+(\hat{R}^2)^2)+\frac{1}{\hat{r}_t}(1-\hat{R}^2)+\hat{R}^2}-\frac{b_t}{2}(1-\hat{R}^2)\right)$$ converges in probability to

$\sqrt{\frac{b_t^2}{4}(1-2R^2+(R^2)^2+\frac{1}{r_t}}-\frac{b_t}{2}(1-R^2)$, $\hat{\beta}_t$ is asymptotically normal with variance

$$[v_{11t}]\left(\sqrt{\frac{b_t^2}{4}(1-2R^2+(R^2)^2+\frac{1}{r_t}}-\frac{b_t}{2}(1-R^2)\right)^2.$$