Identifying Sheepskin Effects in the Returns to Education

Alfonso Flores-Lagunes
Department of Economics
University of Arizona
alfonso@eller.arizona.edu

Audrey Light
Department of Economics
Ohio State University
light.20@osu.edu

April 2004

1. Introduction

A central issue in the economics of education literature is whether credentialed workers (those who receive high school diplomas or college degrees) earn more than observationally equivalent non-credentialed workers. Such "sheepskin effects" are consistent with sorting models of education (Arrow 1973, Spence 1973, Stiglitz 1975, Weiss 1983) in which employers use credentials to identify workers with desirable traits that cannot be directly observed.[1] Most empirical assessments of sheepskin effects are, in fact, aimed at validating the role of schooling as a sorting mechanism (Arkes 1999, Belman and Heywood 1991, 1997, Ferrer and Riddell 2002, Frazis 1993, Hungerford and Solon 1986, Jaeger and Page 1996, Liu and Wong 1982). However, an understanding of sheepskin effects also underlies efforts to measure schooling attainment (Kominski and Siegel 1993, Park 1999), assess the benefits of community college attendance and college transfers (Kane and Rouse 1994, Light and Strayer 2004) and parameterize the relationship between schooling and earnings in a variety of applications (Card 1999).

Despite widespread interest, the magnitude of sheepskin effects has yet to be pinned down. In the earliest empirical studies (Belman and Heywood 1991, Hungerford and Solon 1986), analysts lacked data on degree status (D) and were forced to infer sheepskin effects from nonlinearities in the relationship between highest grade completed (S) and log wages.[2] When data on both S and D became available, researchers found that estimated degree effects conditional on schooling are generally larger than the earlier, indirect estimates (Arkes 1999, Ferrer and Riddell 2002, Frazis 1993, Jaeger and Page 1996, Park 1999). Jaeger and Page (1996) suggest that the indirect estimates are biased downward by measurement error that arises when "expected schooling" is used as a proxy for degree attainment. Their argument does not acknowledge that S and D might be reported with error (a possibility subsequently addressed by Kane, Rouse and Staiger 1999) and it skirts a related issue, which is that estimated sheepskin effects based on identical data are highly sensitive to functional form. Using a model that includes a dummy variable for each year of schooling and each degree level, Jaeger and Page predict a gap in log wages of 0.16 between workers who hold bachelor's degrees and "college dropouts." Using a model that includes a dummy variable for every S-D interaction, they predict

[1] Following Weiss (1995), we use the term "sorting models" to refer to both signaling and screening versions of the models.

[2] Another early study (Taubman and Wales 1973) used data on degrees but not years of schooling.

the same gap in log wages (holding *S* constant at 16) to be 75% higher. This is a particularly stark example of the non-robustness evident throughout the empirical literature on sheepskin effects.

In this study, we ask how to interpret the range of estimated sheepskin effects that arise from different model specifications. We begin with the observation that sheepskin effects are identified because individuals with a given amount of schooling differ in their degree attainment or, stated differently, because *S* varies among individuals within a given degree category. It is important to recognize that the variation needed for identification can represent "signal," "noise," or a combination of both—and that the source of variation determines which parameters are identified and how we should interpret the estimates.

To illustrate the identification issues, suppose we observe variation in *S* among individuals who hold a bachelor's degree. This variation might accurately reflect the underlying behavioral process if, for example, some college students choose full-time (or even part-time) employment at the expense of more rapid progress toward a degree. The variation might also reflect measurement error: data that cross-classify individuals as college graduates with only 12, 13 or 14 years of school should certainly be met with suspicion, and even reports that appear more logical can be error-ridden. Variation in *S* within degree category is essential if we wish to identify *S-D* interactions in a wage model, but it is equally important to consider the source of variation. Under the first scenario, we might predict that the sheepskin effect increases with *S* because *S* is positively correlated with omitted in-school work experience—however, the estimate should be interpreted as a reward for work experience rather than evidence that employers use degrees to screen for unobserved traits. If the variation arises primarily from measurement error, it might pay to restrict the functional form of our wage model rather than rely on "noise" for identification.

We formalize these arguments in the next section. Specifically, we demonstrate how ordinary least squares estimates of sheepskin effects are expected to change as we alter the parameterization of the wage model and the source of variation in the data; we consider different forms of measurement error as well as enrollment behaviors as alternative sources of variation in *S* and *D*. In subsequent sections, we turn to the data to assess the *actual* sensitivity of estimated sheepskin effects to functional form, measurement error, and other factors. We use data from the 1979 National Longitudinal Survey of Youth (NLSY79), which provides data on both *S* and *D* for a large sample of workers along with information on work experience and enrollment patterns. By estimating wage models with and without measures of actual work experience and age at degree recipiency, we attempt to isolate the signaling effect of sheepskins from the confounding effects of factors that explain some of the variation in *S* and *D* but are fully observed by employers. Moreover, because many NLSY79 respondents report their schooling attainment repeatedly over a number of years, we use internal inconsistency in self-reports as a rough indication of measurement error. By reestimating our wage models with *S* and *D* data that are judged to be "clean," we assess the nature of measurement error bias in estimated sheepskin effects and, in particular, the interplay between measurement error bias and functional form.

2. Identification Issues

Consider the following log wage models:

$$\ln W = \sum_j b_j S_j + u \tag{1a}$$

$$\ln W = \sum_j b_j S_j + \sum_k g_k D_k + u \tag{1b}$$

$$\ln W = \sum_j b_j S_j + \sum_k g_k D_k + \sum_{j \neq k} \sum_k d_{jk} S_j D_k + u \tag{1c}$$

where the S_j are dummy variables that equal one if the worker's highest grade completed is *j* (*j*=6,7,…20) and the D_k are dummy variables identifying four mutually exclusive degree categories: high school dropout, high school graduate, college dropout, and college graduate.[3] Individual subscripts and additional covariates (*e.g.,* years of work experience) are suppressed. We begin with these specifications because they have been used in key studies within the "sheepskin" literature: Hungerford and Solon (1987) use model 1a, while Jaeger and Page (1996) estimate all three models.

For comparison to 1a-1c, we also consider the following three specifications:

$$\ln W = bS + u \tag{2a}$$

$$\ln W = bS + \sum_k g_k D_k + u \tag{2b}$$

$$\ln W = bS + \sum_k g_k D_k + \sum_k d_k S \cdot D_k + u \tag{2c}$$

where *S* is now a "continuous" (categorical) measure of highest grade completed. Model 2a is a

[3]When we turn to the data, we include additional categories for holders of associate's degrees and post-bachelor's (graduate) degrees.

useful benchmark, given its prominence in the literature (Mincer 1974, Card 1999), but is obviously ill suited to estimating sheepskin effects. Our primary reason for considering these restricted versions of models 1a-1c is that we wish to cover the spectrum from the simplest model that identifies schooling and degree effects (2b) to the most flexible model (1c) in which each *S-D* combination has its own parameter.

Our goals are to learn why these alternative specifications produce different estimates of sheepskin effects, and which specification is preferred. In principle, the latter objective calls for F-tests to determine which parameter restrictions are acceptable to the data. Rather than rely exclusively on F-tests, however, we ask how the variation in *S* and *D* is generated and how that variation, in turn, is used to identify the parameters of interest.[4] We consider the identification of sheepskin effects under two alternative scenarios. First, we assume our key schooling-related variables (as well as all other explanatory variables) are measured without error. Under this assumption, variation in *S* and *D* reflects interpersonal differences in ability and other factors that lead students to choose different schooling outcomes. Next, we assume that measurement error exists in either *S* or *D*—a scenario in which estimated sheepskin effects could be spurious. By considering two limiting cases where the variation needed for identification represents, alternatively, "signal" and "noise," we gain insights into the more realistic scenario where both sources of variation exist in the data.

A. No measurement error

If *S* and *D* are measured without error, then the variation needed to identify the parameters in models 1a-1c and 2a-2c is the outcome of individuals' schooling decisions. The early "sheepskin" literature does not consider why *S* would vary among individuals with the same *D* or why *D* would vary among individuals with the same *S*. This is unsurprising, for before Jaeger and Page (1996) defined sheepskin effects as the wage gains associated with degrees *conditional on S*, analysts simply looked for an *unconditional* relationship between degrees and wages or "degree years" and wages. According to orthodox sorting models (Arrow 1973, Spence 1973, Stiglitz 1975), individuals with unobserved traits that employers value (ability, determination, *etc.*) receive more schooling because those traits lower the cost of

[4]Cawley *et al.* (2001) and Heckman and Vytlacil (2001) conduct a related exercise. The authors consider models similar to 1a-1c in functional form, but with S and *D* replaced with other variables, and ask how identification is affected by the nature of the data and functional form restrictions.

schooling. As a result, employers can rely on observed components of schooling attainment such as degree status to screen for the desired traits. The theory explains why *D* (or *S*) varies across individuals, and variation in *D* (or *S*) is all that "first generation" empirical analysts (Taubman and Wales 1973, Layard and Psacharopoulos 1974, Hungerford and Solon 1987) need to identify the relationship of interest.[5]

Empirical researchers subsequently began controlling for both *S* and *D* in wage models, but to our knowledge only Weiss (1983) offers a theoretical explanation for why these two variables would vary independently.[6] In his model, individuals choose their optimal level of S and then take a test. Because individuals with high ability (the unobserved trait valued by employers) have a higher probability of passing the test, the test score serves as a useful signal. Given that "passing the test" is equivalent to earning a credential, the model explains why degree holders have varying levels of *S*.

However, Weiss's stylized model does not conform to certain aspects of the U.S. educational system. The "test" that leads to a high school diploma is very different than the "test" that leads to a bachelor's degree. It is necessary to distinguish between degree types and explain why *S* varies *within* each *D* category. The reality appears to be that students who are employed while in school take longer to earn their degrees, as do students who struggle academically. The pattern is slightly different among students in the "dropout" degree categories: in-school employment continues to be positively correlated with time in school, but high ability students tend to drop out later than their low ability counterparts. These behaviors generate variation in *enrollment durations* within each *D* category, but we would see little variation in *S* (highest grade completed) if individuals took explicit account of accumulated credits when reporting S—for example, if students who completed half the credits needed for a bachelor's degree reported themselves as having completed grade 14, regardless of how many years they spent in college. However, it is apparent from the detailed data available in the

[5]Human capital models (Becker 1993, Mincer 1974) provide an alternative explanation for why individuals differ in their schooling attainment and why schooling measures are positively correlated with wages. Chiswick (1973) and Weiss (1995) discuss the difficulties of distinguishing between human capital and sorting hypotheses.

[6]Ferrer and Riddell (2002) refer to wage models that control for *S* only (*e.g*., 1a and 2a) as "human capital models," those that control for *D* only as "credentialist models," and those that control for both S and D (*e.g*., 1b-1c, 2b-2c) as "hybrid models." That is, they recognize that the practice of controlling for both *S* and *D* does not emerge directly from sorting models.

NLSY79 that S measures the number of years spent in school independent of enrollment intensity and progress toward a degree. We document these patterns in section 3.

To see the implications for estimating wage models that control for D and S, we focus on model 1b. Model 1b is a linear model with degree-specific intercepts (fixed effects), so it is clear that b is identified from variation in S within degree category.[7] If S is positively correlated with in-school work experience (or age at school exit) within degree category, then the omission of in-school work experience (or age at school exit) from the model causes the estimate of b to be biased upward. This, in turn, causes the estimate of g_k to be biased downward, given that $\hat{g}_k = \overline{Y}_k - \hat{b}\overline{S}_k$ where $\overline{Y}_k$ and $\overline{S}_k$ are means for group k. In making this assertion we assume, in the spirit of sorting models, that the "true" b is intended to represent the wage effects of schooling *independent* of age, work experience, and other factors that employers can observe directly. The remedy, of course, is to include measures of in-school work experience and related observables in the model. This argument also applies to the estimation of models 1b, 1c and 2c.

We have also suggested that variation in S within D category reflects ability differences, but that S is negatively correlated with ability for degree holders and positively correlated with ability for nondegree holders. This is a departure from the central tenet of sorting models, which is that S and D increase monotonically with ability. Under our scenario, employers should favor degree holders who earn their credential in the shortest time, and favor dropouts who stay in school the longest. This hypothesis can be tested by estimating models 1c and 2c instead of restricted models 1b and 2b and determining whether the "return" to schooling varies across degree categories as predicted. Other analysts (Jaeger and Page 1996, Park 1999) have estimated model 1c and have found that sheepskin effects *do* vary with S; to our knowledge, however, an interpretation of those interaction effects has not previously been offered.

B. Measurement error in *S*

We now assume that at least some of the variation in S represents measurement error. Moreover, we assume that the error is "classical"—that is, we assume $S = S^* + n$, where S is reported schooling, S^* is true schooling and n is a mean-zero, constant variance error term that is uncorrelated with both S^* and the error term in the wage model (u). (We must now replace S

[7] In contrast to models 1c and 2c, which include *S-D* interactions, models 1b and 2b only require variation within *some* degree categories for the schooling coefficients to be identified.

with S^* in models 1a-1c and 2a-2c because each model is intended to express the relationship between *true* schooling and log wages.) The assumption that schooling data exhibit classical measurement error has been questioned by Black, Berger and Scott (2000), Bound and Solon (1999) and Kane, Rouse and Staiger (1999), among others. Because of the categorical nature of S and the fact that it is coded between zero and a top-coded value, these authors suggest that mean-reverting error might be a better assumption. However, in Flores-Lagunes and Light (2003) we use generalized method of moments estimation to fit data to a wide array of error structures, and we find that classical measurement error provides as good a fit as the alternatives.[8]

Under the assumption of classical measurement error in S, we wish to know how measurement error bias varies across model specifications. Beginning with the simplest of our specifications, model 2a, we have the well-known result that

$$\text{plim}\,\hat{\beta} = \frac{\sigma^2_{S^*}}{\sigma^2_S}\beta = \lambda\beta,$$

where $\sigma^2_{S^*}$ and σ^2_S are the variances of S^* and S. In the probability limit, the OLS estimate is only a fraction of the true β, where the fraction (λ) is the ratio of "signal" to "signal plus noise" (also known as the reliability ratio) in reported S.

Model 2b is a degree-specific fixed effects model, so the probability limit of $\hat{\beta}$ becomes

$$\text{plim}\,\hat{\beta}^{FE} = \frac{\sigma^2_{\tilde{S}^*}}{\sigma^2_{\tilde{S}}}\beta = \tilde{\lambda}\beta$$

where $\sigma^2_{\tilde{S}^*}$ and $\sigma^2_{\tilde{S}}$ are the variances of $\tilde{S}^* = S^* - \overline{S}^*_k$ and $\tilde{S} = S - \overline{S}_k$ (deviations from D-specific means). It is straightforward to show that $\tilde{\lambda} < \lambda$ as long as S^* is correlated among individuals with a given D (Ashenfelter and Krueger 1994, Griliches and Hausman 1986). Measurement error bias increases with the correlation in S^* because that correlation determines how much "signal" is lost by the deviations-from-means transformation used to compute $\tilde{\lambda}$. In our application, the S^* are likely to be highly correlated among individuals who attain a given degree

[8] In contemporary U.S. data sources (including the ones used by Flores-Lagunes and Light 2003), very few schooling reports appear in the 0-6 range or at the top-coded value. This may explain why the error structure appears to be classical.

level because of the uniformity of the U.S. educational system. For example, the fact that schools are designed to award high school diplomas upon the completion of grade 12 induces correlation among the “true” schooling level of terminal high school graduates. In addition, individuals who reach a common degree level will be similar in terms of their innate ability, access to funds and preferences—and that similarity would induce correlation in S^* even if schooling and degrees were chosen independently. A downward bias in $\hat{b}$ leads to an upward bias in $\hat{g}_k$ because $\hat{g}_k = \overline{Y}_k - \hat{b}\overline{S}_k$. The widely-reported finding that sheepskin effects identified by model 2b are larger than “degree year” effects implied by model 2a (or by variants of these models) could be entirely due to measurement error in *S*.

Model 2c differs from 2b in that it has *D*-specific slopes as well as *D*-specific intercepts. To obtain estimates for model 2c, we can estimate four models of the form

$$\ln W_k = b_k S_k + g_k + u_k \quad \text{for } k = 1,2,3,4 \tag{3}$$

and restrict the error variances to be equal for all k. (Note that b_k in model 3 is $b + d_k$ in 2c.) For this “unpooled” model we have

$$\text{plim}\, \hat{b}_k = \frac{s^2_{S^*_k}}{s^2_{S_k}} b_k = l_k b_k$$

where $s^2_{S^*_k}$ and $s^2_{S_k}$ are the variances of S^* and S for the N_k observations in group k. It is clear that $l_k < l$ because $s^2_{n_k} = s^2_n$ (reporting errors have the same variance regardless of how and whether we subset the sample) but $s^2_{S_k} < s^2_S$ (reported S has less variation within a *D*-specific subsample than in the overall sample). A comparison of l_k and $\tilde{l}$ is less clear-cut because it depends on whether more variation in *S* is lost by eliminating observations outside the k^{th} degree group (model 2c) or by subtracting *D*-specific means (model 2b). We can illustrate this point by assuming we have only two observations per group—an unrealistic assumption in our application, but one that corresponds to within-family estimation based on samples of twins (Ashenfelter and Krueger 1994). In this case, $l_k < \tilde{l}$ as long as $s^2_{S_k} < s^2_S - \text{cov}(S^*_1, S^*_2)$.[9]

[9]Note that $l = 1 - [s^2_n / s^2_S]$, $l_k = 1 - [s^2_{n_k} / s^2_{S_k}]$ and, in the case of two observations per group,

Unless true schooling (S^*) is “too correlated” within *D*-specific group, measurement error bias is likely to be greater in 2c than in 2b because variation in *S* is significantly smaller within group than in the overall sample. That is, we expect downward bias in the estimated marginal effect of *S* and upward bias in the sheepskin effects (g_k) to increase when we add *S*-*D* interactions to the model.

Next, we ask how measurement error bias is likely to change when we replace S with dummy variables identifying each individual year of schooling. We focus on model 1a, and continue to assume that the error in S is “classical.” Therefore, when we use each reported *S* to create the dummy variables S_j, where $S_j = 1$ if $S = S^* + n = j$, the random component n induces misclassification.[10] Given that the estimated coefficients in model 1a are simply the mean of log wages for individuals in the given *S*-specific category ($\hat{b}_j = \overline{\ln W_j}$), a bias in the estimates will arise due to (a) the inclusion of individuals in group *j* that do not belong there, and (b) the omission of individuals that belong to group *j* but are erroneously included in another category. Assuming that ln*W* increases monotonically with *S*, then the direction of the bias in $\hat{b}_j$ depends on the relative numbers of individuals from categories above and below category *j* that are erroneously placed in category j. This pattern arises even though the random component n is independent of the true (continuous) S^*.

Under this set of assumptions, the estimated coefficient for the bottom (top) *S*-specific dummy will be biased upward (downward) because all misclassified observations belong to individuals whose true schooling would place them in a higher (lower) category and who, therefore, have higher (lower) $\overline{\ln W}$.[11] However, the estimated coefficients for the middle *S*-specific dummies will have a negligible bias: a random portion of observations that belong in this category are excluded (placed in a different category as a result of measurement error), while

$\tilde{l} = 1 - [s_n^2 / (s_S^2 - s_{S^*}^2 r)]$, where r is the correlation in S^* within groups.

[10]Our conclusions regarding model 1a are unchanged if we replace the classical error assumption with the assumption of mean-reverting error.

[11]Top-coded observations are an exception to this pattern. If survey respondents correctly report their true S^* as 21 or higher but are coded as S=20 (the NLSY79 top-code), then they rightfully belong to a higher category than the 20th. In our sample only 1.5% of respondents appear in the highest category (*S*=20), so it is apparent that top-coding is not an important source of bias.

observations that are erroneously placed in this category are randomly drawn from higher and lower categories in a way that leaves $\overline{\ln W}$ roughly equal to its true mean. Because classical measurement error in *S* leads to a nonuniform pattern of misclassification in the S_j dummy variables, we may be able to obtain relatively unbiased estimates of some of the coefficients in models 1a-1c. However, relatively few observations are used to compute $\overline{\ln W}$ for the less common *S* levels—especially in model 1c, where the *S*-specific means are computed *within* each degree category. As a result, a small amount of "nonrandomness" in the misclassifications can produce potentially severe measurement error bias for these models.

C. Measurement Error in *D*

It is more difficult to assess the biases if we assume that *D*, rather than *S*, is measured with error. While *S* can arguably be treated as a continuous measure exhibiting classical measurement error, *D* is a set of dummy variables that classify individuals into one of four mutually exclusive degree levels—*D* is an intrinsically categorical variable. We cannot assume that the errors in D_k are independent of the errors in D_l, nor can we assume that the errors are unrelated to the truth. Aigner (1973), Black *et al*. (2000), Freeman (1984), and Kane *et al.* (1999) discuss cases where a *single* dummy variable is subject to misclassification. They show that measurement error causes the estimated coefficient for the binary variable to be biased downward, as in the case of classical measurement error. Black *et al.* (2000) and Kane *et al.* (1999) demonstrate that the bias continues to be downward when the categorical variable takes on more than two values (see also Griliches 1986). The derivations become extremely complicated for a situation similar to ours, where there are (interrelated) errors in four dummy variables. Rather than attempt such a derivation, we simply note that our qualitative conclusions regarding measurement error in *S* should apply to the case where *D* is measured with error.

The more likely scenario is that both *S* and *D* are measured with error. Kane *et al.* (1999) consider this case, but find that degree status is reported much more accurately than years of schooling. Their finding is another reason that we direct more attention to errors in *S* than to errors in *D* not only in this section, but also in our discussion of the estimates.[12]

[12]Deriving results for the case of two misreported variables is beyond the scope of this paper and, in fact, beyond the scope of most studies. As Greene (2003) notes: "If more than one variable is measured with error, there is very little that can be said."

3. Data

A. Samples

We use data from the 1979 National Longitudinal Survey of Youth (NLSY79) to estimate the wage models described in section 2. The NLSY79 began in 1979 with a sample of 12,686 youth born in 1957-64, and it remains in progress today. Respondents were interviewed annually from 1979 to 1994 and biennially thereafter; 2000 is the last year for which we have data. Additional details on the survey can be found in Center for Human Resource Research (2000).

We obtain a cross-sectional sample of wage observations as follows. First, we select each respondent's last reported wage. This wage refers to a job held in 2000 for many respondents, but it pre-dates 2000 for respondents who were not employed that year or who dropped out of the survey prior to 2000. The resulting sample contains one wage observation for each of 12,098 individuals; the average age is 34 and the age range is 18-44 at the time the wage was earned. Second, we identify the self-reported highest grade completed and degree attainment at the time the wage was earned. If respondents were not enrolled in school when a given interview took place and had not been enrolled since the last interview, they were not asked to report their highest grade completed. Therefore, for respondents who were not "fresh out of school" when their wage was reported, we look across preceding interviews to find the most recent report for S. The receipt of high school diplomas, associate's, bachelor's and graduate degrees, and college enrollment are identified in a similar fashion. Schooling-related variables are missing for a handful of respondents, so at this stage we have a sample size of 12,061.

In addition, we eliminate Hispanic and black respondents from the sample. The sample that remains consists of 7,113 individuals who are almost exclusively white. It is common in the sheepskin literature to focus on white men (Hungerford and Solon 1987) or to estimate separate wage models for different race-sex groups (Belman and Heywood 1991, Jaeger and Page 1996). We eliminate blacks and Hispanics because an examination of race differences in sheepskin effects is beyond the scope of the current study. However, we use a pooled sample of men and women because our NLSY79 sample is considerably smaller than the samples of Current Population Survey (CPS) respondents used in each of these earlier studies. When we estimate models for white men only, the estimated coefficients for the schooling and degree variables are generally very close to what we find with the pooled sample, and in no case do our qualitative conclusions differ across the two samples.[13] We believe the increased precision obtained by doubling the sample size more than offsets the cost of being unable to identify minor differences between sheepskin effects for white men and white women.

In addition to analyzing the entire sample of 7,113 whites, we also examine a subsample in which the schooling and degree variables are judged to be "clean." We eliminate respondents who report highest grade completed (*S*) only once during the survey—*i.e*., respondents who left school prior to the first interview and never returned and, as a result, reported *S* in 1979 only. Among remaining respondents, we require that sequential reports of *S* increment consistently with calendar time: for example, the annual sequence 12,13,14,15 is consistent, but 12,13,15,16 is not. In addition, we require consistency in reported information on degree attainment: individuals who report college attendance must have received a high school diploma or GED at an earlier date, individuals who receive a bachelor's degree must have reported attending college prior to that date, and so forth. These criteria produce a sample of 2,165 whites for whom the *S* and *D* data are not necessarily error-free, but are invariably less error-ridden than the data in the larger sample. By requiring internal consistency among multiple reports, we eliminate a disproportionate share of respondents with low schooling attainment and/or earlier birth years. Nonetheless, we believe this is a useful strategy for assessing the effect of measurement error on the estimates.[14]

B. Variables

Our dependent variable is the log of the average hourly wage divided by the GDP implicit price deflator. Our key explanatory variables are highest grade completed (*S*) and a set of dummy variables (*D*) identifying respondents as having no degree, a high school diploma, college attendance without a degree, an associate's degree, a bachelor's degree, or a graduate (master's, professional or doctorate) degree; the degree categories refer to the highest level attained and are mutually exclusive. Summary statistics for these and other variables appear in

[13]For models 2a-2c, we fail to reject the null hypothesis of equality in b, g_k, and d_k across the two groups (white men versus white men and women) using a 5% significance level. For models 1a-1c, we find some statistically significant (but qualitatively "uninteresting") differences.

[14]Our strategy is less demanding of the data than those requiring validation data (*e.g.*, Freeman 1984), and more flexible than those requiring relatively simple functional forms in order to jointly estimate measurement error and outcome models (Black *et al.* 2000, Flores-Lagunes and Light 2003, Kane *et al.* 1999).

table 1.

Table 2 contains a cross-tabulation of *S* and *D*. It is clear from these distributions that *S* varies considerably within *D* category. For example, almost one out of four respondents with high school diplomas and bachelor's degrees report a schooling level other than the "usual" level of 12 and 16, respectively. Similarly, *D* varies for each level of *S*. Within the *S*=12 subsample, for example, 77% of individuals claim to have a high school diploma but others report every degree level from high school dropout to graduate degree. This variation in the data allows us to identify sheepskin effects—and, in fact, to do so with very flexible models such as 1c and 2c. However, a portion of the variation is likely to be due to measurement error. A primary goal of our analysis is to determine how alternative specifications of the wage model exacerbate or "smooth over" noise in the data.

Each model also includes a dummy variable indicating whether the respondent is male, and a set of calendar year dummies. To be consistent with sheepskin studies based on CPS data (*e.g.*, Belman and Heywood 1991, Hungerford and Solon 1987, Jaeger and Page 1996, Park 1999) we estimate a set of benchmark models in which the only additional variables are potential experience (age-*S*-6) and its square. To follow up on the ideas discussed in section 2A, we then replace potential experience with actual experience (cumulative hours worked from the 20th birthday to the date the wage was earned, divided by 2,000) and its square and cube, plus the age at which the respondent earned his degree or left school. The purpose of these specifications is to control for variables that are correlated with *S* and *D*—and that explain much of the variation in *S* within degree category—but that employers observe directly. In section 2A we suggest controlling for in-school work experience, but after some experimentation we opted to control for cumulative work experience from age 20 onward. We find that estimated schooling coefficients are not sensitive to whether actual experience is divided into the portion obtained in school and the portion obtained after school, although they are highly sensitive to whether in-school experience is controlled for in some fashion (see also Light 1998, 2001). In addition, for many respondents we lack detailed data on work experience gained before age 20.[15]

---

[15]The NLSY79 records employment histories from January 1978 onward, so we can track actual experience from the 20th birthday onward for all respondents born in 1958-64; for those born in 1957, we predict 1977 hours worked. If we were to "start the clock" at age 16 in order to measure in-school work experience for high school dropouts and high school graduates, experience would be left-censored for virtually every respondent.

In section 2a we suggested that, within degree category, *S* is positively correlated with in-school work experience and the age at which the degree is received. Table 3 demonstrates the extent to which these patterns exist in the data. We divide the full sample of 7,113 whites into *D*-specific subsamples, and then further divide them into three groups: reported *S* below the *D*-specific median, *S* equal to the median, and *S* above the median. (We ignore individuals with graduate degrees in table 3 because the subgroups are quite small.) Within each degree group, the mean level of actual work experience and the mean age at school exit increase as S increases. Relatively few of the differences in means are significantly different than zero, but the patterns point to positive correlations nonetheless. In light of these patterns, we believe it is necessary to "net out" the effects of work experience and age to avoid overstating the wage effects of *S* and *D*—especially if sheepskin effects are to be interpreted as evidence of sorting on unobservable traits.

To follow up on the measurement issue raised earlier, for the bachelor's degree subsample we also report mean levels of in-school work experience in table 3. We measure in-school work experience as the average hours worked per week in the year prior to leaving school; this 52-week period is uncensored for all college graduates in the sample. Table 3 shows that average, in-school employment intensity is sharply higher among individuals who complete more than 16 years of school (the median for this group) than it is for other college graduates. Because both cumulative experience and in-school experience increase with *S*, we believe the former variable is a good proxy for the latter.

We also predicted that *S* and ability are positively related within the dropout categories, but negatively related within the "completed degree" categories. We investigate this pattern in table 3, using percentile scores on the Armed Forces Qualifying Test (AFQT) to approximate the unobserved ability that employers might be seeking. Among individuals who leave high school or college without a degree, mean AFQT scores increase dramatically as schooling increases (although, again, with considerable variation and small sample sizes, the differences in means are not always significantly different than zero). We do not see the predicted negative relationship between *S* and test scores among the degree recipients, but the positive relationship appears to be much less pronounced. These disparate patterns suggest that it might be necessary to allow the marginal effect of *S* on log wages to vary across degree levels.

4. Estimates

We estimate a number of wage models that vary in three dimensions. First, as described by models 1a-1c and 2a-2c, we change the functional form of the relationship between schooling, degree status, and log wages. Second, we alter the set of covariates other than years of schooling and degree status. We control for potential experience and its square in some specifications, but we substitute actual experience and age at degree recipiency (or school exit for nondegree holders) in others. Third, we use both the full sample of white workers as well as a smaller sample in which the schooling data are judged to be "clean." In tables 4-6 we report predicted sheepskin effects and marginal effects of schooling for each specification. OLS estimates of the underlying coefficients for a subset of specifications appear in appendix tables A1-2.

We begin by noting the sensitivity to model specification that we alluded to in the introduction. This is readily apparent in table 4, which contains predictions based on models 1a-1c and 2a-2c, using the full sample of white workers and controlling for potential experience. The estimated marginal effect of the 12th year of school ranges from an imprecisely estimated 0.018 (model 1c) to 0.101 (model 2a), while the predicted change in log wage associated with advancing from "some college" to a bachelor's degree ranges from an imprecisely estimated 0.065 (model 1c) to 0.198 (model 2b).[16] This sensitivity in estimated sheepskin effects is reduced when we replace potential experience with actual experience (and add a control for age at which the degree was received) and especially when we switch to the "clean" sample. In table 6, with both those changes in place, the predicted "return" to a bachelor's degree ranges from 0.094 (1b) to 0.189 (2b).

Using F-tests to determine which specification is preferred, we find that the data strongly favor model 1b over the more restrictive model 1a *and* the highly flexible 1c, and we find that model 2c is preferred to 2a and 2b. When we compare 1a to 2a, 1b to 2b and 1c to 2c, we always reject the null hypothesis that the 15 *S*-specific coefficients are equal.[17] If we intended simply to

---

[16]Point estimates for model 1c often differ dramatically from the other estimates, and invariably have large standard errors. With relatively few observations available to compute the mean log wage within each *S-D* cell, these estimates are particularly sensitive to extreme values of the dependent variable as well as measurement error in *S* and *D*.

[17] In comparing 1b and 1c, the p-value for the null hypothesis $(\mathrm{H}_0 : d_{jk} = 0)$ is 0.347. For each of the other comparisons, the p-value is no greater than 0.001. These evaluations are based on models that control for potential experience, but we obtain similar results when we use actual experience and/or "clean" data. We ignore the potential effects of measurement error on the outcome of these F-tests.

find the specification that provides the best fit to the data, we would consider a model that is intermediate to 1b and 2c. That is, we would look for groups of S-specific dummies for which the estimated coefficients can be constrained to be equal, and form interactions between these aggregate S groups and the degree dummies. As stated in section 2, however, we use models 1a-1c and 2a-2c because we wish to span the range from the simplest functional form (2a-2b) to the most flexible (1c). Taking these six specifications as "given," we now ask how to interpret the estimated schooling and degree effects in light of the two potential sources of variation in the data: signal and noise.

To investigate the possibility that the estimated S coefficients partially reflect wage payoffs to work experience gained in school (an upward bias in b that would, in turn, affect the estimated degree effects), we compare estimates from specifications that control for potential experience to those that control for actual experience and age at degree recipiency. The relevant estimates appear in tables 4-5 and A1-A2. When we replace potential experience with more detailed work history variables, the estimated schooling coefficient falls from 0.101 to 0.087 in model 2a, from 0.055 to 0.037 in 2b, and from 0.036 to 0.024 in 2c. The estimated coefficient for S_{16} falls from 0.778 to 0.634 in model 1a and from 0.549 to 0.346 in 1b; the point estimates for most of the remaining S dummies decline in magnitude as well, although using conventional significance levels we do not always reject the null hypothesis that the change is zero. By bringing to bear less information than employers would typically have, we are concerned that the "potential experience" specifications not only overstate the wage payoff to years of schooling, but also understate the degree effects. In model 2b, the estimated coefficients for associate's, bachelor's, and graduate degrees are significantly higher when we replace potential experience with actual experience, and in model 1b all the degree effects increase. With both model 1b and model 2b, for example, the predicted effect of a bachelor's degree relative to "some college" increases by about 0.07 when we replace potential experience with actual experience. If we wish to interpret these sheepskin effects as evidence of labor market signaling, we understate the value of a sheepskin when key observables are omitted from the specification.[18]

Dhrymes (1978) shows that the F-test for the overall significance of the regression in a classical errors-in-variables model is understated, which leads to over-rejection. However, to our knowledge, little can be said about the effect of classical errors in an F-test that compares two different nested models.

[18]If the estimated sheepskin effects instead represent the wage benefits of skills gained in conjunction with each degree level, then we understate the value-added of this skill.

We also investigate the notion that sheepskin effects increase with schooling for high school and college dropouts, but decrease (or fail to increase) with schooling for degree holders. The argument is that high ability individuals remain in school longer before dropping out, while *low* ability individuals take the longest to earn a degree. We find that this prediction is largely supported by the data. Table A1 reveals that the estimated coefficients (standard errors) for the *S-D* interactions for model 2c are -0.001 (.018), 0.053 (.017), 0.061 (.024), -0.008 (.020), and 0.025 (.029) for high school diplomas, some college, associate's degrees, bachelor's degrees and graduate degrees, respectively. In other words, the estimated marginal effect of *S* is significantly higher for college-goers who do not earn a bachelor's degree than for others, although there is no evidence that "late" high school dropouts earn more than "early" high school dropouts. When we switch to model 1c, we fail to reject the null hypothesis that the coefficients for the *S-D* interactions are zero (see footnote 17). Nonetheless, the very noisy point estimates for this model are entirely consistent with our argument: the predicted log wage for a high school dropout with 11 years of school exceeds that of his counterpart with S=9 by 0.14; for college dropouts, the predicted log wage increases by 0.15 when we increase S from 13 to 15 while for high school graduates, the predicted log wage falls by -0.20 when we increase S from 12 to 13. These patterns suggest that the more flexible models 1c and 2c capture differential signaling across degree categories.

Next, we ask whether the estimates differ across specifications in a manner consistent with measurement error. In section 2, we argued that the estimated coefficient for S is expected to be increasingly downward biased due to measurement error as we move from model 2a to 2b to 2c. Although the same intuition applies to models 1a-1c, we argued further that misclassification (and, therefore, measurement error bias) is likely to be less severe for intermediate schooling years than for the top and bottom categories. In fact, we find that most estimated S coefficients decrease when we add degree dummies and they decrease further when we add *S-D* interactions. For example, the estimated coefficient (standard error) for S decreases from 0.101 (.004) to 0.055 (.006) to 0.036 (.013) as we move from model 2a to 2b to 2c. The estimated coefficient for S_6 (the bottom S category) is an imprecisely estimated -0.26 in both 1a and 1b, while the estimated coefficient for S_{14} falls from 0.483 (.042) to 0.338 (.053) and the estimated coefficient for S_{20} (the top category) falls from 1.029 (.068) to 0.656 (.091).

We can further assess measurement error bias by comparing the estimates in table 5 to the estimates in table 6, which are based on the sample in which reports for *S* and *D* are judged to be "clean." If *S* is measured with error, then the estimated *S* coefficients in table 5 are downward biased and the estimated *D* coefficients are upward biased. When we switch to clean data, we indeed find that for models 2a-2c the estimated *S* coefficients *increase* and the estimated *D* coefficients *decrease*. In model 2b, for example, the estimated schooling coefficient increases from 0.037 (.006) to 0.047 (.012) when we switch to the clean sample, while the estimated coefficient for a bachelor's degree decreases fom 0.407 (.048) to 0.313 (.096) and the estimated coefficient for "some college" decreases from 0.144 (.032) to 0.123 (.073) (table A1). This pattern is entirely consistent with the assumption of classical error in *S* and no error in *D*.

The expected patterns are more complicated for models 1a and 1b because misclassification should lead to downward (upward) bias for the estimated coefficient for the top (bottom) *S* category, and misclassification should be less severe for intermediate values of *S*. In switching from the full sample to the "clean" sample, we find that *every* estimated *S* coefficient for models 1a-1b declines in value, although the changes are not always statistically significant. Focusing on model 1a we find, for example, that the estimated coefficient for S_{14} declines from 0.361 to 0.110 while the estimated coefficient for S_{15} falls from 0.451 to 0.348 and the estimate coefficient for S_{20} (the top category) falls from 0.917 to 0.824 (table A2). It does not appear that misclassification of observations in the intermediate categories is random, nor does it appear that the top category is biased downward by misclassification in the full sample. Instead, we conclude that the very small samples used to compute a coefficient for each *S*-specific dummy (and especially for each *S*-*D*-specific dummy in model 1c) make these estimates highly sensitive to nonrandom errors.

5. Conclusions

In the 1990s, analyst began identifying sheepskin effects by exploiting data sources that include measures of workers' highest grade completed *and* their degree attainment. Specifically, they estimated log wage models that control for highest grade completed (*S*), highest degree attained (*D*) and, in some cases, interactions between *S* and *D*. The wage effect of *D*, conditional on *S*, is referred to as a sheepskin effect. It is often interpreted as evidence that employers use degrees to screen for workers with high ability and other desired traits that cannot be directly observed.

While data sources that measure both *S* and *D* are unquestionably valuable, it is important to ask why these two measures would vary independently. Sheepskin effects are identified because individuals with a given amount of schooling differ in their degree attainment, or because individuals with a given degree complete different years of *S*. This variation might be the outcome of individuals' schooling decisions, or it might be purely due to measurement error in *S* or *D*. In this study, we ask how to interpret estimated sheepskin effects when the variation arises from these alternative scenarios.

Within a given *D* category, we demonstrate that individuals who gain more S also gain more work experience while in school and are older at school exit. Employers invariably observe these characteristics as easily as they observe degree attainment, so it is a mistake to identify sheepskin effects from models that rely on variation in *S* and *D*, but do not control for these related factors. We find that sheepskin effects are over-stated (by 7% or so for the "value added" of a bachelor's degree) in models that ignore in-school work experience and age at school exit. We also find that increased *S* might be a *positive* signal to employers hiring dropouts (especially individuals who attend college without earning a bachelor's degree) because high ability dropouts stay in school longer, but a *negative* signal to employers hiring degree recipients. Models that include *S-D* interactions are able to identify these differential "returns" to degrees.

At the same time, we find that estimated sheepskin effects might be *over-stated* as a result of measurement error in *S*. Earlier studies (*e.g.*, Jaeger and Page 1996) correctly highlight the fact that self-reported degree information eliminates the need to infer degree attainment from data on highest grade completed. However, measurement error in *S* can lead to downward bias in estimated coefficients for *S* and upward bias in estimated coefficients for *D*. Our estimates based on log wage models that control for "continuous" *S* strongly support this measurement error argument. It is much more difficult to assess the role of measurement error in models that replace *S* with dummy variables for each highest grade completed. Our estimates for those models are not entirely consistent with simple assumptions about the nature of the reporting errors. Additional analysis of measurement error in wage models that include a large number of dummy variables is at the top of our agenda for future research.

References

Aigner, Dennis J. "Regression With a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics* 1 (March 1973): 49-59.

Arkes, Jeremy. "What Do Educational Credentials Signal and Why Do Employers Value Credentials?" *Economics of Education Review* 18 (February 1999): 133-141.

Arrow, Kenneth J. "Higher Education as a Filter." *Journal of Public Economics* 2 (July 1973): 193-216.

Ashenfelter, Orley and Alan Krueger. "Estimates of the Economic Returns to Schooling from a New Sample of Twins." *American Economic Review* 84 (December 1994): 1157-1173.

Becker, Gary. *Human Capital*, 3rd edition. Chicago: University of Chicago Press, 1993.

Belman, Dale and John S. Heywood. "Sheepskin Effects in the Returns to Education: An Examination of Women and Minorities." *Review of Economics and Statistics* 73 (November 1991): 720-24.

——— and ———. "Sheepskin Effects by Cohort: Implications of Job Matching in a Signaling Model." *Oxford Economic Pap*ers 49 (October 1997): 623-637.

Black, Dan A., Mark C. Berger and Frank A. Scott. "Bounding Parameter Estimates with Nonclassical Measurement Error." *Journal of the American Statistical Association* 95 (September 2000): 739-748.

Bound, John and Gary Solon, "Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling," *Economics of Education Review* 18 (April 1999): 169-182.

Card, David. "The Causal Effect of Education on Earnings." In Orley Ashenfelter and David Card (editors), *Handbook of Labor Economics,* Volume 3. Amsterdam: Elsevier Science B.V., 1999.

Cawley, John, James Heckman, and Edward Vytlacil. "Three Observations on Wages and Measured Cognitive Ability." *Labour Economics* 8 (September 2001): 419-442.

Center for Human Resource Research. *NLSY79 User's Guide*. 2000. Columbus, OH: The Ohio State University, CHRR NLS User Services.

Chiswick, Barry. "Schooling, Screening, and Income." In Lewis Solmon and Paul Taubman (editors), *Does College Matter?* New York: Academic Press, 1973.

Dhrymes, Phoebus. *Introductory Econometrics*. New York: Springer-Verlag, 1978.

Ferrer, Ana M. and Craig W. Riddell. "The Role of Credentials in the Canadian Labour Market." *Canadian Journal of Economics* 35 (November 2002): 879-905.

Flores-Lagunes, Alfonso and Audrey Light. "Measurement Error in Schooling: Evidence from Samples of Siblings and Identical Twins." Unpublished paper, October 2003.

Frazis, Harvey. "Selection Bias and the Degree Effect." *Journal of Human Resources* 28 (Summer 1993): 538-554.

Freeman, Richard B. "Longitudinal Analyses of the Effects of Trade Unions." *Journal of Labor Economics* 2 (January 1984): 1-26.

Greene, William H. *Econometric Analysis* (5th edition). Upper Saddle River, NJ: Prentice Hall, 2003.

Griliches, Zvi. "Economic Data Issues." In Zvi Griliches and Michael D. Intriligator (editors), *Handbook of Econometrics,* Volume 3. Amsterdam: Elsevier Science B.V., 1986.

——— and Jerry A. Hausman. "Errors in Variables in Panel Data." *Journal of Econometrics* 31 (February 1986): 93-118.

Heckman, James J. and Edward Vytlacil. "Identifying the Role of Cognitive Ability in Explaining the Level of Change in the Returns to Schooling." *Review of Economics and Statistics* 83 (February 2001): 1-12.

Hungerford, Thomas and Gary Solon. "Sheepskin Effects in the Returns to Education." *Review of Economics and Statistics* 69 (February 1987): 175-77.

Jaeger, David A. and Marianne E. Page. "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education." *Review of Economics and Statistics* 78 (November 1996): 733-740.

Kane, Thomas J. and Cecilia Elena Rouse. "Labor-Market Returns to Two- and Four-Year College." *American Economic Review* 85 (June 1995): 600-614.

———, ——— and Douglas Staiger. ""Estimating Returns to Schooling When Schooling is Misreported." NBER Working Paper 7235, July 1999.

Kominski, Robert and Paul M. Siegel. "Measuring Education in the Current Population Survey." *Monthly Labor Review* 116 (September 1993): 34-38.

Layard, Richard and George Psacharopoulos. "The Screening Hypothesis and the Returns to Education." *Journal of Political Economy* 82 (September/October 1974): 985-998.

Light, Audrey. "Estimating Returns to Schooling: When Does the Career Begin?" *Economics of Education Review* 17 (February 1998): 31-45.

———. "In-School Work Experience and the Returns to Schooling." *Journal of Labor Economics* 19 (January 2001): 65-93.

——— and Wayne Strayer. "Who Receives the College Wage Premium? Assessing the Labor Market Returns to Degrees and College Transfer Patterns." *Journal of Human Resources 3*9 (forthcoming Summer 2004).

Liu, Pak-Wai and Yue-Chim Wong. "Educational Screening by Certificates: An Empirical Test." *Economic Inquiry* 20 (January 1982): 72-83.

Mincer, Jacob. *Schooling, Experience, and Earnings.* New York: Columbia University Press, 1974.

Park, Jin-Heum. "Estimation of Sheepskin Effects Using the Old and the New Measures of Educational Attainment in the Current Population Survey." *Economics Letters* 62 (February 1999): 237-240.

Spence, Michael. "Job Market Signaling." *Quarterly Journal of Economics* 87 (August 1973): 355-374.

Stiglitz, Joseph E. "The Theory of 'Screening,' Education, and the Distribution of Income." *American Economic Review* 65 (June 1975): 283-300.

Taubman, Paul J. and Terence J. Wales. "Higher Education, Mental Ability, and Screening." *Journal of Political Economy* 81 (January/February 1973): 28-55.

Weiss, Andrew. "A Sorting-cum-Learning Model of Education." *Journal of Political Economy* 91 (June 1983): 420-442.

———. "Human Capital vs. Signaling Explanations of Wages." *Journal of Economic Perspectives* 9 (Fall 1995): 133-154.

Table 1: Summary Statistics for Selected Variables

	All whites		Clean sample[a]	
Variable	Mean	S.D.	Mean	S.D.
Ln(average hourly wage)	2.43	.68	2.52	.72
Highest grade completed (*S*)	13.01	2.65	13.59	2.61
Degree (*D*)				
None	.12		.07	
High school diploma	.37		.41	
Some college	.22		.16	
Associate's degree	.07		.06	
Bachelor's degree	.16		.24	
Graduate degree	.06		.06	
Potential experience (Age-S-6)	15.46	6.16	14.51	5.69
Actual experience[b]	12.00	7.43	12.32	6.84
Age received degree or left school	20.81	4.77	20.54	3.74
1 if male	.51		.51	
Number of individuals	7,113		2,165	

[a]Excludes individuals with (a) only one report of S, (b) multiple reports of S that vary inconsistently with calendar time, or (c) inconsistent degree reports.

[b]Hours worked from 20th birthday to date wage was earned, divided by 2000.

Table 2: Cross-Tabulations of Highest Grade Completed by Highest Degree Received (all whites)

Highest grade completed	None	High school	Some college	Assoc. degree	Bach. degree	Graduate degree	All degree levels	Sample size
0-8	83.6	13.7	1.2		1.5			262
	25.8	**1.4**	**0.2**		**0.4**		**3.7**	
9-11	55.3	41.8	2.5	0.3	0.1			1,117
	72.9	**17.6**	**1.8**	**0.6**	**0.1**		**15.7**	
12	0.2	76.6	22.0	1.1	0.2	0.0		2,696
	0.5	**77.7**	**38.2**	**5.8**	**0.4**	**0.3**	**37.9**	
13-15	0.4	6.5	60.1	29.7	3.2	0.2		1,383
	0.6	**3.4**	**53.6**	**82.0**	**3.8**	**0.8**	**19.4**	
16	0.1		7.5	3.5	88.3	0.7		1,007
	0.1		**4.8**	**7.0**	**77.0**	**1.7**	**14.2**	
17-20	0.2		3.2	3.6	32.7	60.3		648
	0.1		**1.4**	**4.6**	**18.4**	**97.3**	**9.1**	
All grades	11.9	37.4	21.8	7.0	16.2	5.7		
Sample size	848	2,658	1,550	500	1,155	402		7,113

Note: Nonbold-face numbers are percents of row totals and bold-face numbers are percents of column totals.

Table 3: Summary Statistics for Selected Characteristics, by Degree and Highest Grade Completed (all whites)

Variable	Highest grade completed relative to median for degree group: $S < S^{median}$		$S = S^{median}$		$S > S^{median}$	
	Mean	S.D	Mean	S.D.	Mean	S.D.
High school dropout						
Age left school	17.13	4.14	17.86	3.56	19.29	4.13
Actual work experience	7.86	6.53	8.84	7.03	9.10	7.21
Percentile score on AFQT	15.63	15.47	18.18	15.53	23.99	18.65
High school dipoma						
Age received degree	18.18	5.39	18.69	2.14	18.25	4.94
Actual work experience	10.09	7.07	11.50	7.87	12.43	8.15
Percentile score on AFQT	33.30	18.75	43.64	21.71	46.03	22.24
Some college						
Age left school	18.31	1.17	18.67	2.40	18.80	2.13
Actual work experience	12.62	10.85	15.35	12.53	16.42	13.46
Percentile score on AFQT	49.37	20.94	54.89	23.12	65.09	22.98
Associate's degree						
Age received degree	21.99	4.65	23.91	5.76	25.49	5.99
Actual work experience	13.39	5.64	13.93	6.66	15.20	6.34
Percentile score on AFQT	49.50	23.68	59.95	21.20	60.51	22.13
Bachelor's degree						
Age received degree	23.17	3.83	24.41	4.04	24.82	4.02
Actual work experience	13.69	6.94	14.49	6.07	14.75	5.64
In-school experience[a]	18.01	13.14	18.01	10.40	25.43	7.39
Percentile score on AFQT	63.02	22.47	73.93	19.13	73.75	18.47

[a]Average hours worked per week during the last year of college.

Note: Each degree-specific sample is disaggregated according to whether reported S is below, equal to, or above the degree-specific median. The medians are 11,12,13,14 and 16 for each successive degree.

Table 4: Predicted Wage Effects of Schooling and Degrees
(all whites, controlling for potential experience)

Predicted marginal effect of increasing from:	Holding constant:[a]	Specification					
		2a	1a	2b	1b	2c	1c
S=10 to S=11	HS dropout	.101 (.004)	.003 (.040)	.055 (.006)	-.005 (.040)	.036 (.013)	-.027 (.056)
S=11 to S=12	HS diploma	same	.091 (.029)	same	.027 (.033)	.035 (.012)	.018 (.039)
S=12 to S=13	HS diploma	same	.160 (.027)	same	.106 (.030)	same	.088 (.069)
S=13 to S=14	Some college	same	.077 (.033)	same	.076 (.036)	.089 (.011)	.050 (.042)
S=14 to S=15	Associate's degree	same	.083 (.043)	same	.070 (.040)	.096 (.021)	-.029 (.068)
S=15 to S=16	Bachelor's degree	same	.212 (.042)	same	.141 (.052)	.028 (.015)	-.028 (.101)
S=16 to S=17	Bachelor's degree	same	.055 (.045)	same	.019 (.046)	same	-.002 (.055)
HS dropout to HS diploma	S=12			.056 (.026)	.105 (.032)	.101 (.041)	-.511 (.281)
HS diploma to some college	S=14			.120 (.020)	.087 (.023)	.199 (.034)	.054 (.133)
Some college to Associate's degree	S=14			.015 (.030)	-.009 (.033)	-.028 (.031)	.025 (.047)
Some college to Bachelor's degree	S=16			.198 (.029)	.084 (.043)	.107 (.039)	.065 (.069)

[a]Degree and schooling levels used to compute predictions for specifications 1c and 2c.

Note: Standard errors are in parentheses. The sample consists of 7,113 white men and women. "Same" means the estimate is constrained to equal the estimate in the preceding row.

Table 5: Predicted Wage Effects of Schooling and Degrees
(all whites, controlling for actual experience and age at which degree was earned)

Predicted marginal effect of increasing from:	Holding constant:[a]	Specification					
		2a	1a	2b	1b	2c	1c
S=10 to S=11	HS dropout	.087 (.003)	-.027 (.038)	.037 (.006)	-.033 (.038)	.024 (.013)	-.043 (.054)
S=11 to S=12	HS diploma	same	.041 (.029)	Same	-.039 (.032)	-.005 (.011)	-.038 (.039)
S=12 to S=13	HS diploma	same	.140 (.026)	Same	.086 (.029)	Same	.039 (.067)
S=13 to S=14	Some college	same	.084 (.032)	Same	.073 (.034)	.068 (.010)	.038 (.041)
S=14 to S=15	Associate's degree	same	.089 (.042)	Same	.071 (.043)	.083 (.020)	-.007 (.065)
S=15 to S=16	Bachelor's degree	same	.183 (.040)	Same	.082 (.050)	.024 (.015)	-.058 (.097)
S=16 to S=17	Bachelor's degree	same	.057 (.043)	Same	.018 (.044)	Same	.002 (.053)
HS dropout to HS diploma	S=12			.031 (.026)	.120 (.031)	.050 (.040)	-.402 (.272)
HS diploma to some college	S=14			.113 (.019)	.083 (.022)	.238 (.033)	.038 (.128)
Some college to Associate's degree	S=14			.070 (.031)	.033 (.033)	.032 (.032)	.081 (.047)
Some college to Bachelor's degree	S=16			.263 (.030)	.160 (.043)	.178 (.039)	.158 (.067)

[a]Degree and schooling levels used to compute predictions for specifications 1c and 2c.

Note: Standard errors are in parentheses. The sample consists of 7,113 white men and women. "Same" means the estimate is constrained to equal the estimate in the preceding row.

Table 6: Predicted Wage Effects of Schooling and Degrees
("clean" sample, controlling for actual experience and age at which degree was earned)

Predicted marginal effect of increasing from:	Holding constant:[a]	Specification					
		2a	1a	2b	1b	2c	1c
S=10 to S=11	HS dropout	.096 (.006)	.032 (.082)	.047 (.012)	.018 (.083)	.015 (.033)	-.073 (.124)
S=11 to S=12	HS diploma	same	.018 (.056)	same	-.046 (.062)	-.033 (.031)	-.065 (.071)
S=12 to S=13	HS diploma	same	.158 (.056)	same	.041 (.074)	same	.116 (.217)
S=13 to S=14	Some college	same	.025 (.071)	same	.034 (.074)	.085 (.022)	-.001 (.083)
S=14 to S=15	Associate's degree	same	.239 (.090)	same	.224 (.090)	.098 (.031)	.096 (.136)
S=15 to S=16	Bachelor's degree	same	.077 (.081)	same	.009 (.094)	.026 (.025)	-.174 (.192)
S=16 to S=17	Bachelor's degree	same	.088 (.072)	same	.082 (.073)	same	.009 (.082)
HS dropout to HS diploma	S=12			-.022 (.056)	.127 (.072)	.027 (.088)	-.385 (.333)
HS diploma to some college	S=14			.145 (.043)	.140 (.058)	.338 (.077)	.068 (.408)
Some college to Associate's degree	S=14			.003 (.062)	-.016 (.065)	-.030 (.064)	.073 (.099)
Some college to Bachelor's degree	S=16			.189 (.056)	.094 (.073)	.110 (.070)	.097 (.098)

[a]Degree and schooling levels used to compute predictions for specifications 1c and 2c.

Note: Standard errors are in parentheses. The sample consists of 2,165 white men and women who report S consistently as described in the text and the note to table 1. "Same" means the estimate is constrained to equal the estimate in the preceding row.

Table A1: OLS Estimates for Selected Model 2 Specifications

	2a (table 4)		**2b (table 4)**		**2a (table 5)**		**2b (table 5)**	
Variable	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Constant	.547	.096	.987	.107	.714	.048	1.335	.072
S	.101	.004	.055	.006	.087	.003	.037	.006
1 if HS diploma			.056	.026			.031	.026
1 if some college			.177	.033			.144	.032
1 if assoc. degree			.191	.043			.214	.043
1 if bach. degree			.374	.047			.407	.048
1 if graduate degree			.111	.035			.179	.035
Potential exper. (PX)	.035	.008	.032	.008				
$PX^2/100$	-.087	.021	-.075	.021				
Actual exper. (AX)					.049	.009	.053	.009
$AX^2/100$					-.009	.068	-.046	.068
$AX^3/1000$					-.021	.015	-.012	.016
Age received degree					-.001	.002	-.008	.002
1 if male	.319	.014	.319	.013	.220	.014	.218	.014
Observations	7,113		7,113		7,113		7,113	
Root MSE	.5616		.561		.547		.541	

	2a (table 6)		**2b (table 6)**		**2c (table 4)**		**2c (table 5)**	
Variable	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Constant	.605	.101	1.297	.166	1.152	.154	1.467	.125
S	.096	.006	.047	.012	.036	.013	.024	.013
1 if HS diploma			-.022	.056	.110	.187	.398	.186
1 if some college			.123	.073	-.445	.188	-.391	.182
1 if assoc. degree			.127	.090	-.579	.319	-.560	.310
1 if bach. degree			.313	.096	.632	.278	.493	.269
1 if graduate degree			.174	.069	-.296	.511	-.334	.494
S*HS diploma					-.001	.018	-.029	.018
S*some college					.053	.017	.044	.016
S*assoc. degree					.061	.024	.059	.024
S*bach. degree					-.008	.020	.000	.019
S*graduate degree					.025	.029	.030	.028
Potential exper. (PX)					.035	.008		
PX^2					-.081	.021		
Actual exper. (AX)	.029	.019	.032	.019			.053	.009
$AX^2/10$	.027	.015	.024	.015			-.047	.067
$AX^3/100$	-.010	.004	-.009	.004			-.012	.016
Age received degree	.002	.004	-.006	.005			-.009	.002
1 if male	.205	.027	.210	.027	.318	.013	.215	.014
Observations	2,165		2,165		7,113		7,113	
Root MSE	.575		.571		.559		.540	

Note: Each specification also includes calendar year dummies.

Table A2: OLS Estimates for Selected Model 1 Specifications

	1a (table 4)		1b (table 4)		1a (table 5)		1b (table 5)		1a (table 6)		1b(table 6)	
Variable	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Constant	1.448	.078	1.423	.078	1.602	.055	1.655	.057	1.837	.132	1.884	.136
S=6	-.026	.109	-.028	.110	.049	.106	.034	.106	-.364	.414	-.313	.413
S=7	.024	.080	.030	.080	.034	.077	.037	.077	-.337	.270	-.397	.270
S=8	.009	.054	.022	.054	-.012	.052	-.000	.053	-.119	.146	-.103	.146
S=10	.152	.043	.134	.043	.123	.042	.106	.042	-.124	.106	-.146	.107
S=11	.155	.043	.129	.043	.096	.041	.073	.042	-.092	.099	-.128	.101
S=12	.246	.035	.156	.042	.137	.034	.034	.040	-.073	.087	-.175	.100
S=13	.406	.042	.262	.050	.277	.040	.120	.048	.085	.099	-.133	.122
S=14	.483	.042	.338	.053	.361	.039	.193	.050	.110	.098	-.100	.121
S=15	.566	.052	.408	.061	.451	.049	.264	.058	.348	.115	.125	.137
S=16	.778	.042	.549	.063	.634	.037	.346	.059	.425	.091	.135	.127
S=17	.833	.057	.568	.073	.691	.052	.364	.068	.513	.111	.216	.140
S=18	.936	.055	.580	.079	.811	.050	.392	.073	.618	.114	.245	.151
S=19	1.048	.073	.705	.092	.915	.067	.506	.085	.620	.127	.244	.162
S=20	1.029	.068	.656	.091	.917	.061	.478	.083	.824	.122	.418	.163
1 if HS dipl.			.105	.032			.120	.031			.127	.071
1 if some coll.			.192	.038			.203	.037			.266	.088
1 if associate			.173	.049			.236	.050			.250	.104
1 if bachelors			.276	.056			.363	.057			.360	.111
1 if graduate			.168	.050			.191	.049			.152	.092
PX	.035	.008	.034	.008								
$PX^2/100$	-.085	.021	-.080	.021								
AX					.053	.009	.052	.009	.035	.019	.035	.019
$AX^2/100$					-.044	.068	-.043	.067	.221	.148	.219	.148
$AX^3/1000$					-.003	.016	-.012	.016	-.090	.035	-.086	.035
Age recd. deg.					-.005	.002	-.009	.002	-.003	.004	-.008	.005
1 if male	.317	.014	.318	.013	.212	.034	.214	.014	.198	.027	.205	.027
Observations	7,113		7,113		7,113		7,113		2,165		2,165	
Root MSE	.561		.559		.543		.540		.572		.570	

Note: Each specification also includes calendar year dummies.