Review of Economic Studies (2001) **68**, 665–691 0034-6527/01/00280667$02.00

The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?

CHRISTOPHER R. TABER
Northwestern University

First version received February 1998; *final version accepted August* 2000 (*Eds.*)

The changes in the distribution of earnings during the 1980s have been studied extensively. The two most striking characteristics of the decade are (a) a large increase in the college/high school wage gap, and (b) a substantial rise in the variance of wage residuals. While this second phenomenon is typically implicitly attributed to an increase in the demand for unobserved skill, most work in this area fails to acknowledge that this same increase in demand for unobserved skill could drive the evolution of the measured college premium. In its simplest form, if higher ability individuals are more likely to attend college, then the increase in the college wage premium may be due to a increase in the relative demand for high ability workers rather than an increase in the demand for skills accumulated in college. This paper develops and estimates a dynamic programming selection model in order to investigate the plausibility of this explanation. The results are highly suggestive that an increase in the demand for unobserved ability could play a major role in the growing college premium.

1. INTRODUCTION

The changes in the distribution of earnings during the 1980s have been studied extensively. The two most striking characteristics of the decade are (a) a large increase in the college/high school wage gap, and (b) a substantial rise in the variance of wage residuals. While this second phenomenon is typically implicitly attributed to an increase in the demand for unobserved skill,[1] most work in this area fails to acknowledge that this same increase in demand for unobserved skill could drive the evolution of the measured college premium. In its simplest form, if higher ability individuals are more likely to attend college, then the increase in the college wage premium may be due to an increase in the relative demand for high ability workers rather than an increase in the demand for skills accumulated in college. This paper develops and estimates a dynamic programming selection model in order to investigate the plausibility of this explanation. The results are highly suggestive that an increase in the demand for unobserved ability could play a major role in the growing college premium.

There are at least two reasons to be interested in this topic. First, for policy makers intent on helping the American workforce adjust to the changing wage structure this distinction is crucial. If the value added of attending college truly rises, they may consider policies aimed at subsidizing the skills learned in college. However, if it results from an increase in the payoff to other types of skills, then policies focused on skill acquisition earlier in life would be more appropriate. Second, it will also aid in understanding the

1. See for example Juhn, Murphy and Pierce (1993).

nature of the change in the demand for skills. For example, recent papers by Acemoglu (1999) and Beaudry and Green (1998), argue not only that there has been an increase in the demand for college skill, but that the increase in college skill actually drove the changes in the college wage gap. This explanation seems much less plausible if the rise in the college premium is due to a rise in the demand for ability rather than a rise in the demand for college educated workers. Another line suggests that the rise of computers may have changed the workplace (Krueger, 1993; Autor, Katz and Krueger, 1999). It is not clear which skills taught in college should be complementary with computer use, but it seems quite feasible that the type of individual who chooses to attend college might be more likely to excel at computer work. For example individuals who enjoy learning may be more adept at adjusting to the new technology and may be more likely to attend college.

2. BACKGROUND AND RELATED WORK

To make the problem explicit, suppose earnings are determined by the following regression function similar to that described in Griliches (1977, 1979)

$$Y_t = \alpha_t(s) + \gamma_t \theta + \varepsilon_t, \tag{1}$$

where Y_t is some measure of earnings, s is years of schooling, and θ is ability.[2]

For two levels of schooling $s_1 > s_2$, the observed wage differential can be written as

$$E[Y_t | S = s_1] - E[Y_t | S = s_2] = \alpha_t(s_1) - \alpha_t(s_2) + \gamma_t(E[\theta | S = s_1] - E[\theta | S = s_2]).$$

At a given point in time the observed differential is composed of both the payoff to schooling, $\alpha_t(s_1) - \alpha_t(s_2)$, and "ability bias", $\gamma_t(E[\theta | S = s_1] - E[\theta | S = s_2])$. A positive differential may result either because there are earnings gains from schooling, $(\alpha_t(s_1) - \alpha_t(s_2) > 0)$, or because higher ability individuals earn more and tend to stay in school longer ($\gamma_t > 0$ and $E[\theta | S = s_1] > E[\theta | S = s_2]$). Similarly, the increase in the college premium could be due to one of three possibilities; (1) an increase in $\alpha_t(s_1) - \alpha_t(s_2)$ (change in the payoff to college), (2) an increase in γ_t (change in the payoff to unobserved ability), or (3) an increase in $(E[\theta | S = s_1] - E[\theta | S = s_2])$ (changes in the ability differential between college and high school workers). Since the increase in the college premium occurs within cohorts, changes in the relative quality of workers cannot be the major factor in the changing premium. It is impossible to identify separately time effects such as γ_t, from age effects, and cohort effects such as $(E[\theta | S = s_1] - E[\theta | S = s_2])$. Some assumption is necessary for identification. Since the cohorts in this analysis are at most seven years apart, and since there were not big changes in college matriculation during the period in which they attended, it seems unlikely that cohort effects could be driving the results with this data set. In what follows, I ignore cohort effects by assuming that I am distinguishing between the payoff to college and the payoff to ability, but it should be kept in mind that it is possible that cohort effects could be playing some role as well.

While almost no work has examined the change to unobserved ability, a number of researchers have studied the trend in the return to cognitive ability as measured by scores on standardized tests. Murnane, Willet and Levy (1993) combine the National Longitudinal Study of the High School Class of 1972 and the High School and Beyond data sets to compare wages in 1978 with wages in 1986. They find a large increase in the coefficient on the test score and that including the test score substantially reduces the increase in the

2. The concept of ability is intentionally left abstract, but could manifest itself in observables such as test scores or as intangibles which are unobservable to an empiricist like motivation or positive personality traits.

college premium. Bishop (1991); Blackburn and Neumark (1993); Heckman and Vytlacil (2000*b*); and Cawley *et al.* (2000) employ AFQT scores in the NLSY. The results vary substantially across these studies. Heckman and Vytlacil (2000*b*) provide the most thorough study using a large number of specifications and demonstrate the sensitivity of the results to the maintained assumptions. They find that it is very difficult to separate the two effects because education is so highly correlated with test scores.

An alternative to estimating returns to unobserved skill is to examine patterns of wage dispersion and decompose the patterns into various components as in Card and Lemieux (1993). Chay and Lee (1999) employ an approach of this type. They find evidence of an increase in the return to skill, but, in contrast to the results below, find that it cannot be large enough to account for the full increase in the return to schooling. Their approach is very different from mine so it is difficult to precisely reconcile the results. The major departure comes from the specification of unobserved ability. Chay and Lee (1997) impose one factor of unobserved heterogeneity while I allow for two. With more than one factor, selection bias can increase substantially without imposing a large increase in the variance of wages as is required in a one factor model. They assume that the difference in the ability differential between college and high school workers does not change over time. It may change over time not because of differences in who attends college, but because the combination of skills that constitutes ability that has changed over time in a manner that favours college educated workers.

Bartel and Sicherman (1999) do not look at changes over time explicitly, but look across sectors. They find that the return to college is higher in sectors with higher technological growth, but that "the education premium associated with technological change is the result of a greater demand for the innate ability or other unobserved characteristics of more educated workers". These results are quite similar to mine, but come from a very different approach.

This paper extends the classic work of Willis and Rosen (1979) in several directions. First, the focus of this work is on changes in the return to schooling rather than the level. I use a two factor model which makes more efficient use of the panel data at my disposal in identifying this effect. Second, the Willis and Rosen model is somewhat peculiar in that it is concerned with educational selection, but conditions on high school graduation. I weaken this assumption by modeling high school and college selection together using a dynamic programming model. I treat the decision to attend college as a two stage procedure; students first graduate from high school before they enroll in college. Along with Keane and Wolpin (1997) (whose model is very different from mine) this constitutes one of the first attempts to control for sample selection bias with a dynamic programming model. Other papers which account for the uncertainty in schooling decisions include Comay, Melnik and Pollachek (1973); Altonji (1993); Cameron and Heckman (1998*a*); and Buchinsky and Leslie (1999). This paper also builds on work on the changing wage structure (see Levy and Murnane, 1992, or Autor and Katz, 2000 for surveys) and to the literature on the returns to schooling (see Willis, 1986, or Card, 2000 for surveys).

3. THE DATA

My data come from a representative sample of males from the National Longitudinal Survey of Youth (NLSY), a panel data set begun in 1979 with youth aged 14 to 22. This research was conducted using white male civilians from the cross-sectional sample only. The survey is conducted annually and respondents are questioned on a large range of topics, including schooling, earnings, and family background.

I assign each worker to one of three levels of schooling attainment; if his highest grade of schooling completed is less than 12 years then he is a dropout, if he completes exactly 12 years of schooling then he is a high school graduate, and if he completes at least one year of college then he is a college attender.[3]

My earnings measure is the log of weekly earnings which I construct by dividing annual earnings by weeks worked. This earnings measure was top-coded at $4000 per week and bottom coded at $10 per week. Experience is coded as Age-Education-6 and when I construct experience squared I divide by 100.

In 1980, 94% of the respondents in the NLSY were administered the Armed Services Vocational Aptitude Test (ASVAB) which consists of ten standardized tests which are used by the armed forces to assess a variety of skills. Four of these tests are combined to form the Armed Forces Qualifying Test (AFQT) which is used as a criterion for admission into the armed forces. Both the AFQT score and some of the scores on other sections of the test will be used as measures of skill that may influence both earnings and schooling decisions. I normalize the AFQT score by year of birth by subtracting the mean score for an individual's birth year cohort from his raw score.

TABLE 1

NLSY white males (High school graduates only)

Variable	Mean	Standard deviation
Non time varying variables:		
College	0·544	0·498
Years of schooling	13·857	2·131
Highest grade father	12·677	3·175
Highest grade mother	12·333	2·233
Number of siblings	2·824	1·824
General science score	0·672	4·135
Mathematics knowledge score	1·108	6·119
Auto and shop score	0·422	4·322
Word knowledge score	1·242	5·729
AFQT score	4·10	16·297
Age January 1, 1978	16·390	2·208
Time varying variables:		
Weekly wage	385·398	278·651
Age	25·257	3·093
Experience	5·823	3·233
Experience squared	0·444	0·419
Local unemployment rate	2·950	1·069

In Table 1, I present sample means and variances for the key variables. Since the main wage specifications will focus on individuals who have completed high school, I restrict the sample to those individuals. The mean of the test scores is not zero in this table because they were normalized on the whole NLSY sample, rather than this subsample.

4. PRELIMINARY RESULTS

The first empirical question I address is the extent to which the increase in the college premium can be explained by an increase in the payoff to observed ability as measured

3. Individuals who drop out of school and later receive a General Equivalency Degree (GED) are treated as dropouts. This assumption is supported by Cameron and Heckman (1993) who show that the earnings of GEDs are closer to dropouts than to high school graduates. A GED is given to high school dropouts who pass an exam and is supposed to be equivalent to a high school diploma. Approximately 10% of this sample of white males receive a GED.

by the AFQT score. Several recent papers have shown that conditioning on AFQT can have powerful results (see Herrnstein and Murray, 1994; or Neal and Johnson, 1996). Since there has been a substantial amount of previous work with this data on the subject, I address this question only briefly.

In Table 2 and in those that follow, I divide the data into three time periods, 1982–1984, 1985–1987, and 1988–1990. With the panel data, I potentially observe nine wage

TABLE 2

Preliminary regression results, NLSY white males. Dependent variable: Log weekly wage (Huber standard errors in parentheses) †

Variable	(1)	(2)	(3)	Ages 24–26
1982–1984:				
College	0·370 (0·050)	0·276 (0·053)	0·307 (0·062)	0·288 (0·094)
AFQT score	—	0·577 (0·135)	—	—
General science score	—	—	–0·007 (0·008)	–0·021 (0·015)
Gen. sci. × college	—	—	0·014 (0·014)	0·028 (0·024)
Mathematics knowledge score	—	—	0·013 (0·005)	0·008 (0·009)
Math. × college	—	—	0·003 (0·009)	0·010 (0·014)
Auto and shop score	—	—	0·014 (0·006)	0·018 (0·012)
Auto × college	—	—	0·005 (0·011)	0·001 (0·017)
Word knowledge score	—	—	–0·000 (0·005)	0·007 (0·010)
Word × college	—	—	–0·020 (0·012)	–0·037 (0·020)
Highest grade college	0·104 (0·021)	0·084 (0·021)	0·091 (0·024)	0·079 (0·032)
1985–1987:				
College	0·447 (0·037)	0·350 (0·040)	0·347 (0·046)	0·318 (0·072)
AFQT score	—	0·590 (0·119)	—	—
General science score	—	—	–0·009 (0·009)	–0·010 (0·013)
Gen. sci. × college	—	—	0·008 (0·013)	–0·002 (0·019)
Mathematics knowledge score	—	—	0·016 (0·005)	0·020 (0·007)
Math. × college	—	—	0·003 (0·008)	0·004 (0·010)
Auto and shop score	—	—	0·015 (0·006)	0·021 (0·009)
Auto × college	—	—	–0·005 (0·009)	–0·001 (0·013)
Word knowledge score	—	—	–0·002 (0·006)	–0·000 (0·009)
Word × college	—	—	–0·003 (0·009)	–0·006 (0·013)
Highest grade college	0·113 (0·013)	0·094 (0·014)	0·086 (0·015)	0·083 (0·026)
1988–1990:				
College	0·485 (0·035)	0·402 (0·038)	0·416 (0·041)	0·412 (0·064)
AFQT score	—	0·523 (0·099)	—	—
General science score	—	—	–0·016 (0·008)	0·003 (0·011)
Gen sci. × college	—	—	0·016 (0·011)	0·000 (0·014)
Mathematics knowledge score	—	—	0·018 (0·005)	0·015 (0·007)
Math. × college	—	—	–0·002 (0·007)	–0·000 (0·009)
Auto and shop score	—	—	0·012 (0·005)	0·000 (0·007)
Auto × college	—	—	–0·004 (0·008)	0·005 (0·009)
Word knowledge score	—	—	0·002 (0·005)	–0·005 (0·007)
Word × college	—	—	–0·011 (0·007)	0·005 (0·010)
Highest grade college	0·123 (0·013)	0·107 (0·103)	0·106 (0·014)	0·093 (0·023)
Experience	0·103 (0·010)	0·106 (0·010)	0·101 (0·010)	0·076 (0·028)
Experience squared	–0·375 (0·070)	–0·385 (0·069)	–0·356 (0·069)	–0·144 (0·248)
Local unemployment rate	–0·063 (0·010)	–0·060 (0·010)	–0·065 (0·010)	–0·066 (0·014)
Age January 1, 1978	0·008 (0·007)	0·008 (0·007)	0·009 (0·007)	0·009 (0·010)
Period intercepts	Yes	Yes	Yes	Yes
R^2	0·189	0·205	0·211	0·105
Number of observations	8831	8831	8831	3192
Number of individuals	1527	1527	1527	1308

† Standard errors are calculated for OLS results allowing for heteroskedasticity in individual variance covariance matrices.

observations per individual (three in each period). Of primary interest is the coefficient on "College", which is a dummy variable for college attendance. I control for years of college, but normalize it to zero for people with four years of college. Thus, the dummy variable indicates college attendance, but the coefficient can be interpreted as the return to four years of college.

The first column reports regression results that do not control for test scores. In this specification the coefficient on college rises substantially from 0·370, to 0·447, to 0·485. I next include the AFQT score in the regression to see how much of this increase can be attributed to changes in the payoff to cognitive ability. While including the test score does have a sizeable impact on the levels of the coefficient, it has essentially no effect on the trend. The coefficient of the test score itself also does not rise monotonically, but stays fairly constant across the periods. In the third specification, rather than employ simply the AFQT score, I control for four different components of the ASVAB test that are meant to measure different types of skill. These are also interacted with college attendance. Once again I find no evidence that an increase in the payoff to test scores contributes to the increase in the payoff to college.[4]

One problem in interpreting these results is distinguishing the change in the return to schooling over time from age effects. That is, given that the NLSY panel ages over time, it is possible that the college/time interaction results from the fact that the return to college increase with age rather than over time. I find essentially no evidence of age effects in schooling returns. I address this problem in the fourth column of Table 2 where I condition on individuals age 24–26. While the estimates become noisier, if anything there is stronger evidence of an increasing payoff to college over time. I also estimated a model allowing for interactions between college and both a linear time trend and linear age effect.[5] The college/age interaction in this specification is –0·005 with a standard error of 0·005 while the college/time interaction is 0·044 with a standard error of 0·012. These results strongly suggest that the college/time effect that is the focus of this paper represents a time effect rather than an age effect.

The lack of evidence of a change in the return to *observable* ability does not imply that there has not been a substantial increase in the return to *unobserved* ability. First, there are many aspects to ability and test scores may not be a good proxy for the skill that has increased. Second, failure to reject a null hypothesis does not mean the effect is zero. In results available from the author by request, I follow the methodology of Altonji, Elder and Taber (2000) and show that these results on AFQT scores are sufficiently noisy so that even if the trend in unobserved ability is similar to the trend in observed ability, one cannot reject the null hypothesis that a change in the return to unobserved ability can fully account for the changing college premium.

In order to get some preliminary notion of whether the increase in the measured return to schooling is due to an increase in the "true" return, I present some instrumental variable estimates in Table 3. When I instrument with a variable that is orthogonal to unobserved ability, unobserved ability bias is no longer problematic. Following Willis and Rosen (1978), I exclude family background (father's education, mother's education, number of siblings) from the wage equation. Although the focus of this paper is on the return

4. Taber (1998) provides a much more thorough examination of this question and shows that this result is robust in a substantial number of specifications.

5. I also tried simply interacting schooling with age and including it in the main specification of Table 2. This actually leads to a fairly large *negative* coefficient on the college/age interaction. However given the collinearity between age and time it is hard to interpret this regression in which there is a linear age effect but nonlinear time effect.

TABLE 3

Instrumental variable results using family background as instrument. Dependent variable: log weekly wages (Huber standard errors in parentheses)†*

Variable	OLS	IV‡	OLS	IV‡
1982–1984:				
Schooling	0·079 (0·015)	0·159 (0·069)	0·054 (0·019)	0·100 (0·090)
General science score	−0·018 (0·009)	−0·026 (0·011)	−0·015 (0·010)	−0·018 (0·013)
Mathematics knowledge score	0·008 (0·006)	−0·004 (0·012)	0·012 (0·006)	0·006 (0·015)
Auto and shop score	0·023 (0·007)	0·031 (0·010)	0·020 (0·007)	0·024 (0·013)
Word knowledge score	0·009 (0·007)	0·004 (0·008)	0·000 (0·007)	0·000 (0·007)
1985–1987:				
Schooling	0·083 (0·010)	0·125 (0·034)	0·078 (0·012)	0·121 (0·040)
General science score	−0·008 (0·007)	−0·012 (0·008)	−0·008 (0·008)	−0·012 (0·009)
Mathematics knowledge score	0·020 (0·004)	0·013 (0·007)	0·019 (0·004)	0·012 (0·008)
Auto and shop score	0·012 (0·004)	0·016 (0·006)	0·012 (0·005)	0·016 (0·007)
Word knowledge score	0·001 (0·004)	−0·002 (0·004)	−0·003 (0·005)	−0·003 (0·005)
1988–1990:				
Schooling	0·107 (0·009)	0·133 (0·025)	0·102 (0·010)	0·115 (0·028)
General science score	−0·006 (0·005)	−0·009 (0·006)	−0·009 (0·006)	−0·011 (0·006)
Mathematics knowledge score	0·017 (0·003)	0·012 (0·005)	0·018 (0·003)	0·015 (0·006)
Auto and shop score	0·011 (0·003)	0·014 (0·004)	0·010 (0·004)	0·012 (0·005)
Word knowledge score	0·001 (0·003)	−0·001 (0·003)	−0·002 (0·003)	−0·003 (0·004)
Experience	0·085 (0·012)	0·093 (0·021)	0·069 (0·015)	0·085 (0·022)
Experience squared	−0·219 (0·070)	−0·290 (0·133)	−0·130 (0·097)	−0·252 (0·144)
Local unemployment rate	−0·061 (0·010)	−0·059 (0·011)	−0·061 (0·011)	−0·057 (0·012)
Age January 1, 1978	−0·001 (0·007)	−0·001 (0·007)	0·001 (0·007)	0·001 (0·007)
Period intercepts	Yes	Yes	Yes	Yes
Dropouts included	Yes	Yes	No	No
Number of observations	8394	8394	7113	7113
Number of individuals	1794	1794	1498	1498

* I treat schooling, experience and experience squared as endogenous and instrument excluding parent's education, age and age squared interacted with time periods.
† Standard errors are calculated for OLS results allowing for heteroskedasticity in individual variance covariance matrices.
‡ This table also uses individuals over 22 so that the number of panel observations per individual will not strongly depend on schooling (this does substantially affect the results).

to college attendance rather than on years of schooling, in this case I follow the tradition of the IV literature on schooling by using linear functional forms. The first two columns of estimates include all levels of schooling (as is typical in this literature) and the second two exclude dropouts (as is done in the rest of this paper). Both samples lead to the same basic result, the increase in the schooling premium is eliminated. These regressions suggest that the full increase in the schooling premium over the 1980s was due to an increase in the return to unobserved ability.

There are two reasons to be skeptical of this finding. The first is that one may question the validity of these exclusion restrictions. It seems reasonable that parent's ability could be correlated with their children's ability, and thus parent's education would be positively correlated with their children's ability. However, in results available on request, I tested for the validity of these instruments assuming that the presence of a college and local labour market variables are valid instruments following Cameron and Taber (2000). I performed the tests both using the specification above (on a smaller sample) and using Cameron and Taber's (2000) specification. I did not reject the validity of these exclusion restrictions in any case even at the 20% level. When I exclude the "valid" instruments and include parental education in the wage equation the variables consistently show up as small negative numbers that are statistically insignificant. On the other hand parent's education has a strong effect on schooling. For example, controlling for AFQT scores almost eliminates family income as a determinant of college, but these family background variables are much more robust (see *e.g.* Cameron and Heckman, 1998*b*). This effect of parent's education on schooling could operate either through the tastes for college or through borrowing constraints. While these exclusion restrictions may not be ideal, I see no clear superior alternative for this question. The hope is that they are at least less biased than ordinary least squares (OLS).

More importantly, it is difficult to see why this problem would bias the results towards eliminating the increase in the return to schooling. Consider the simplified specification

$$Y_{it} = \beta_t S_i + \gamma_t \theta_i + \varepsilon_{it},$$

where as above Y_{it} is earnings, θ_i is unobservable ability, ε_{it} is an iid error term, and S_i represents years of schooling. OLS essentially identifies

$$\beta_t^{\text{OLS}} \approx \frac{E(Y_{it} S_i)}{E(S_i^2)}$$
$$\approx \beta_t + \gamma_t \frac{E(\theta_i S_i)}{E(S_i^2)}.$$

Allowing PE_i to represent parent's education of individual i, the instrumental variable estimator essentially identifies

$$\beta_t^{\text{IV}} \approx \frac{E(Y_{it} PE_i)}{E(S_i PE_i)}$$
$$\approx \beta_t + \gamma_t \frac{E(\theta_i PE_i)}{E(S_i PE_i)}.$$

The results in Table 3 suggest that β_t^{IV} was unchanged over time. The standard assumption about the changing wage structure is that the increase in the measured return to college is causal and not due to an increase in the return to unobserved ability. In the notation above this would correspond to β_t increasing over time and γ_t fixed. Within this framework, this explanation cannot be reconciled with values of β_t^{IV} that do not change over time. If parental income were positively correlated with ability there would be an upward bias in the level of the estimates, but not the slope. Two explanations that are consistent with Table 3 are either that (a) the true causal return β_t is fairly flat over time, γ_t is increasing, and the bias in the IV estimate is much smaller than in OLS, or (b) the true causal return β_t is increasing over time, but the IV bias is *much larger* than in OLS and γ_t is *decreasing* over time fast enough so that this bias overwhelms the causal effect. Given that earnings inequality is rising over time, the first explanation seems more plausible.[6]

The second main reason to be skeptical about the result above is the size of the standard errors. Even though the increase in the education return disappears, we cannot

6. An alternative possibility is that $E(\theta_i PE_i)/E(S_i PE_i)$ is falling over time. Given that these individuals are at most 7 years apart in year of birth and that we are using panel data following the same cohorts over time it is difficult to believe this effect could dominate.

reject the original OLS estimates. While the results may not convince one that the measured increase in the return to college is due to increasing ability, they should at least suggest that it may be possible.

I also follow Willis and Rosen's (1978) framework by using the inverse Mills ratio (Heckman, 1979) to estimate the change in the return to college. In the basic model, high school graduates attend college if

$$X'\beta + \theta + \omega > 0,$$

where X is observable, θ is mean zero normally distributed unobserved ability, and ω is normally distributed and independent of wages.[7] I assume that college attenders and high school graduates work in different sectors indexed by school level s. I define high school as $s = 2$, and college as $s = 3$. Wages in the two sectors take the form

$$W_{ts} = Z'_{ts}\gamma_{ts} + \delta_{ts}\theta + \varepsilon_{ts},$$

for $t = 1, 2, 3$ and $s = 2, 3$. Thus the return to ability (both observed and unobserved) may vary across time and sector. This model resembles what Amemiya (1985) calls the type 3 tobit model. Note that

$$\begin{aligned} E(W_{t3}|s = 3, Z_{t3}, X) &= Z'_{t3}\gamma_{t3} + E(\delta_{t3}\theta|s = 3, X) \\ &= Z'_{t3}\gamma_{t3} + \delta_{t3}\sigma_\theta\lambda_c(X'\beta), \\ E(W_{t2}|s = 2, Z_{t2}, X) &= Z'_{t2}\gamma_{t2} + E(\delta_{t2}\theta|s = 2, X) \\ &= Z'_{t2}\gamma_{t2} + \delta_{t2}\sigma_\theta\lambda_h(X'\beta), \end{aligned}$$

where $\lambda_c(X'\beta) \equiv \phi(X'\beta)/\Phi(X'\beta)$, $\lambda_h(X'\beta) = -\phi(X'\beta)/(1 - \Phi(X'\beta))$, and σ^2_θ is the variance of θ.

Average derivatives from the probit model are presented in Table 4. A potential problem that arises with this specification is that the ASVAB scores may induce cohort effects because the test was administered to all respondents in 1980. At that time they were at different ages and schooling levels. In particular, if attending college improves AFQT scores, and since students in the older cohorts may have enrolled in college prior to taking the test, one would expect AFQT to have more predictive power for the older cohorts than for the younger ones. However, at least with a probit specification not reported here, this does not appear to be the case. There is no systematic relationship between birth cohort and the coefficient on AFQT scores. Furthermore, the results are easier to interpret when year of birth and AFQT are not interacted.

The results from the second stage are presented in Table 5. In the first column I repeat the basic specification from Table 2 that does not include the selection correction term. In the second column, I include the Mill's ratio terms and interact them with time. In column three I allow the coefficient to vary with time, but restrict it to be the same in each sector. Column (4) reports a specification that includes a full array of cohort dummies rather than just a linear term. All three specifications yield the same result. Inclusion of the Mill's ratio terms not only eliminates the trend in the college coefficient, but actually leads to a substantial negative trend in the schooling return. This result should be interpreted with much caution. The individual coefficients on the Mills terms are neither individually nor jointly significant. The results in column (2) are statistically indistinguishable from the results in column (1). Even though one cannot reject the joint hypothesis that

7. With normal error terms the distinction between θ and ω can be made without placing any restrictions on the model.

TABLE 4

Selection equation for attend college, average derivatives from probit.† *High school graduates only, NLSY white males (Standard errors in parentheses)*

Variable	(1)	(2)
Highest grade mother	0·022 (0·008)	0·021 (0·008)
Highest grade father	0·033 (0·006)	0·034 (0·006)
Number of siblings	−0·025 (0·008)	−0·026 (0·008)
General science score	0·020 (0·006)	0·021 (0·006)
Word knowledge score	0·010 (0·004)	0·010 (0·004)
Mathematics knowledge score	0·036 (0·003)	0·036 (0·003)
Auto and shop score	−0·033 (0·004)	−0·033 (0·004)
Age January 1, 1978	0·010 (0·007)	—
Age 14 on January 1, 1978	—	−0·011 (0·060)
Age 15 on January 1, 1978	—	0·001 (0·056)
Age 16 on January 1, 1978	—	−0·019 (0·058)
Age 17 on January 1, 1978	—	−0·066 (0·062)
Age 18 on January 1, 1978	—	−0·066 (0·062)
Age 19 on January 1, 1978	—	0·135 (0·056)
Age 20 on January 1, 1978	—	0·065 (0·060)
Number of individuals	1527	

† These results give the derivatives of the predicted probability at the means of the covariates.

the coefficients are zero, there is some evidence that the return to unobserved ability rises over time in specification (3). The test of the null hypothesis that the coefficient on the Mill's ratio is the same in the first and third periods yields a t-statistic of 1·86. The basic finding that controlling for unobserved ability eliminates the college trend is also consistent with the IV results above and with the results from the structural model below. Once again, while these results are not conclusive, they show evidence of an increase in the payoff to unobserved ability that could be an important component of the rising college wage gap.

While the standard errors are large, the point estimates for both the IV procedure and the selection model point to an increasing payoff of the return to unobserved ability. Both of these methods rely strongly on the exclusion restrictions. IV estimation requires them. With linear indices, identification of the selection model does not formally require an exclusion restriction. However, for nonparametric identification an exclusion restriction is required and in practice is often very important. This is the case here as well. Without the exclusion restrictions the results yield unreasonable estimates of the returns to schooling (large negative effects). However, an exclusion restriction is not sufficient for identification; the other aspects of the IV and selection model are important as well. Heckman (1990, 1997) discusses the differences and shows that IV and selection methods estimate different parameters. Since it is not guaranteed that these methods will produce similar results, the fact that the trend in the college premium is eliminated in both cases strengthens the evidence. I will briefly discuss the intuition for the difference between the two in the context of this problem. I draw on this discussion later when contrasting identification in the structural model with these simpler strategies.

The IV strategy makes use of the exclusion restrictions without conditioning on college. Abstracting from other covariates and considering a single instrument X, define the linear model as

$$S_i = \beta X_i + u_i,$$

$$w_{it} = \alpha_t S_i + \delta_t \theta_i + \varepsilon_{it},$$

TABLE 5

Results from two step sample selection model, NLSY white males. Dependent variable: log weekly wage (Huber standard errors in parentheses) †

Variable	(1)	(2)	(3)	(4)
1982–1984:				
College	0·307 (0·062)	0·428 (0·167)	0·432 (0·159)	0·412 (0·151)
Intercept	5·241 (0·108)	5·175 (0·135)	5·177 (0·126)	5·262 (0·101)
Highest grade college	0·091 (0·024)	0·089 (0·024)	0·089 (0·024)	0·089 (0·024)
General science score	−0·007 (0·008)	−0·009 (0·009)	−0·009 (0·009)	−0·010 (0·009)
Gen. sci. × college	0·014 (0·014)	−0·001 (0·005)	−0·001 (0·005)	−0·002 (0·005)
Mathematics knowledge score	0·013 (0·005)	0·009 (0·008)	0·009 (0·007)	0·008 (0·008)
Math. × college	0·003 (0·009)	0·017 (0·008)	0·017 (0·008)	0·018 (0·008)
Auto and shop score	0·014 (0·006)	0·014 (0·016)	0·014 (0·014)	0·016 (0·015)
Auto × college	0·006 (0·011)	−0·021 (0·013)	−0·021 (0·012)	−0·020 (0·013)
Word knowledge score	−0·000 (0·005)	0·004 (0·013)	0·004 (0·009)	0·007 (0·011)
Word × college	−0·020 (0·012)	0·005 (0·013)	0·005 (0·011)	0·002 (0·012)
λ_c	—	−0·071 (0·163)	−0·080 (0·093)	−0·032 (0·093)
λ_h	—	−0·084 (0·113)	−0·080 (0·093)	−0·104 (0·093)
1985–1987:				
College	0·347 (0·046)	0·366 (0·142)	0·379 (0·145)	0·283 (0·136)
Intercept	5·236 (0·101)	5·124 (0·134)	5·220 (0·118)	5·266 (0·108)
Highest grade college	0·086 (0·015)	0·087 (0·015)	0·085 (0·015)	0·087 (0·015)
General science score	−0·009 (0·009)	−0·013 (0·011)	−0·010 (0·010)	−0·012 (0·010)
Gen. sci. × college	0·008 (0·013)	−0·005 (0·006)	−0·003 (0·006)	−0·004 (0·006)
Mathematics knowledge score	0·016 (0·005)	0·009 (0·009)	0·015 (0·007)	0·012 (0·008)
Math. × college	0·003 (0·008)	0·020 (0·009)	0·015 (0·007)	0·018 (0·008)
Auto and shop score	0·015 (0·006)	0·015 (0·015)	0·008 (0·013)	0·015 (0·015)
Auto × college	−0·005 (0·009)	0·002 (0·009)	−0·003 (0·009)	0·002 (0·009)
Word knowledge score	−0·003 (0·006)	0·015 (0·011)	0·003 (0·007)	0·016 (0·011)
Word × college	−0·003 (0·009)	−0·015 (0·011)	−0·005 (0·009)	−0·015 (0·011)
λ_c	—	0·114 (0·115)	−0·021 (0·087)	0·169 (0·087)
λ_h	—	−0·145 (0·127)	−0·021 (0·087)	−0·093 (0·087)
1988–1990:				
College	0·416 (0·041)	0·265 (0·123)	0·260 (0·122)	0·287 (0·121)
Intercept	5·224 (0·094)	5·278 (0·121)	5·300 (0·105)	5·361 (0·099)
Highest grade college	0·106 (0·014)	0·107 (0·014)	0·107 (0·014)	0·107 (0·014)
General science score	−0·016 (0·008)	−0·014 (0·008)	−0·014 (0·008)	−0·015 (0·008)
Gen. sci. × college	0·016 (0·011)	0·003 (0·005)	0·004 (0·005)	0·003 (0·005)
Mathematics knowledge score	0·018 (0·005)	0·022 (0·007)	0·023 (0·006)	0·021 (0·007)
Math. × college	−0·002 (0·007)	0·009 (0·007)	0·008 (0·006)	0·010 (0·007)
Auto and shop score	0·012 (0·005)	0·016 (0·012)	0·015 (0·012)	0·018 (0·012)
Auto × college	−0·004 (0·008)	−0·009 (0·008)	−0·010 (0·007)	−0·009 (0·008)
Word knowledge score	0·002 (0·005)	0·001 (0·010)	−0·001 (0·007)	0·002 (0·010)
Word × college	−0·011 (0·007)	−0·005 (0·009)	−0·003 (0·008)	−0·007 (0·009)
λ_c	—	0·120 (0·095)	0·098 (0·073)	0·117 (0·073)
λ_h	—	0·069 (0·114)	0·098 (0·073)	0·044 (0·073)
Experience	0·102 (0·010)	0·101 (0·010)	0·101 (0·010)	0·103 (0·010)
Experience squared	−0·356 (0·069)	−0·353 (0·070)	−0·347 (0·069)	−0·366 (0·070)
Local unemployment rate	−0·065 (0·010)	−0·065 (0·010)	−0·065 (0·010)	−0·065 (0·010)
Age January 1, 1978	0·009 (0·007)	0·009 (0·007)	0·009 (0·007)	—
Cohort dummies	No	No	No	Yes‡
Number of observations	8831	8831	8831	8831
Number of individuals	1527	1527	1527	1527

† Standard errors are calculated for OLS results allowing for heteroskedasticity in individual variance covariance matrices and correcting for estimation of Mills ratio in first step.
‡ In this case I use cohort dummies in both the first and second stage, in the others only the cohort trend is included in both stages.

where X_i is uncorrelated with the error terms, but u_i and thus S_i is correlated with θ_i. Writing the model in terms of its reduced form yields

$$w_{it} = \alpha_t \beta X_i + \alpha_t u_i + \delta_t \theta_i + \varepsilon_{it}.$$

The source for identification in this reduced form specification is transparent. The result in Table 3 arises from the fact that the reduced form coefficient on parent's education does not change over time.

The selection model makes use of the exclusion restrictions conditionally in the sense that I include college in the wage equation and interact it with the Mill's ratio. To gain an intuition into the selection model, again abstract from other covariances but assume that S_i is binary so that

$$S_i = 1(X_i'\beta + \theta_i + u_i \geqq 0),$$

$$w_{it} = \begin{cases} \alpha_{1t} + \delta_{1t}\theta_i + \varepsilon_{i1t}, & S_i = 1, \\ \alpha_{0t} + \delta_{0t}\theta_i + \varepsilon_{i0t}, & S_i = 0, \end{cases}$$

where θ_i is unobserved ability with $E(\theta_i | X_i) = 0$. Again suppose that X_i is uncorrelated with the error terms and for expositional simplicity that $(u_i, \varepsilon_{i1t}, \varepsilon_{i0t})$ is independent of θ_i. Identification comes from the conditional expectation

$$E(w_{it} | X_i, S_i) = \alpha_{0t} + (\alpha_{1t} - \alpha_{0t})S_i + \delta_{0t}(1 - S_i)E(\theta_i | X_i'\beta + \theta_i + u_i < 0, X_i) + \delta_{1t} S_i E(\theta_i | X_i'\beta + \theta_i + u_i \geqq 0, X_i).$$

In contrast to the IV model, there are two coefficients that pick up the change in the return to unobserved skill δ_{0t} and δ_{1t}. Consider identification of δ_{1t}. When $X_i'\beta$ is very large and $S_i = 1$, the term $E(\theta_i | X_i'\beta + \theta_i + u_i \geqq 0, X_i)$ is close to zero since the unconditional expectation of θ_i is zero. Similarly, this term is large when $X_i'\beta$ is very small. Thus the variable is large for individuals who attend college, but based on their observables, look as if they would be unlikely to attend ($X_i'\beta$ is small). These are individuals who are assigned large amounts of "unobserved ability". In particular, individuals from disadvantaged family backgrounds that attend college would be treated as if they possessed large amounts of unobserved ability. An increase in the coefficient on this interacted Mill's ratio over time indicates that these individuals who look as if they should not go to college, but have anyway, have done well during the 1980s. Using the analogous argument, the Mill's term for high school graduates is $E(\theta_i | X_i'\beta + \theta_i + u_i < 0, X_i)$. This term is close to zero when $X_i'\beta$ is very small and becomes more negative as $X_i'\beta$ increases. Thus it is negative and large in absolute value for individuals who based on their observables look as if they should attend college, but do not. This corresponds to individuals from very privileged family backgrounds that do not attend college. These individuals are assigned low amounts of "unobserved ability". The increasing coefficient on the high school Mill's ratio indicates that they did poorly in the 1980s.

5. ECONOMETRIC SPECIFICATION OF STRUCTURAL MODEL

5.1. *The model*

I assume that there are three levels of schooling corresponding to High School Dropouts ($s = 1$), High School Graduates ($s = 2$), and College attenders ($s = 3$). Students make schooling choices to maximize the expected value of lifetime utility. For each individual, I define V_s to be the lifetime value function that he would achieve if he entered the labour force with schooling attainment s. At the beginning of time period 1 the student is enrolled in high school and decides whether to enter the labour force immediately ($s = 1$) or to stay in school and receive his high school diploma. If he stays in school, he graduates at the end of period 1, and at the beginning of period 2 he again decides whether to enter the labour force immediately ($s = 2$) or to attend college ($s = 3$). With this timing in mind and defining E_t as the student's expectation at the beginning of time period t, I assume that

$$E_1[V_1] = X_1'\beta_1 + \alpha_1' E_1[\theta] + \omega_1, \tag{2}$$

$$E_2[V_2] = X_2'\beta_2 + \alpha_2' E_2[\theta] + \omega_2, \tag{3}$$

$$E_2[V_3] = X_3'\beta_3 + \alpha_3' E_2[\theta] + \omega_3, \tag{4}$$

where α_1, α_2 and α_3 are vectors allowing the value of school to differ for individuals with different sets of unobserved skills and E_t denotes expectations based on information at time t. The random vector θ of skills enters the wage equation as well as potentially influencing schooling directly. Agents learn about the level of their skills as they age. The additional error terms ω_1, ω_2 and ω_3 are independent of the wage equation. Note that for both high school graduation and college attendance, the expectation is defined in terms of information in period 2 because the decision between these two options is made at that point.

The student first decides whether to drop out of high school and then decides whether to complete college. The decision to drop out of high school depends on the total expected value of completing high school which takes the form

$$E_1[\tilde{V}_2] \equiv E_1[\max\{X_2'\beta_2 + \alpha_2' E_2[\theta] + \omega_2, X_3'\beta_3 + \alpha_3' E_2[\theta] + \omega_3\}]$$

$$= E_1[X_2'\beta_2 + \alpha_2' E_2[\theta] + \omega_2] + E_1[\max\{X_3'\beta_3 - X_2'\beta_2 + (\alpha_3 - \alpha_2)' E_2[\theta] + \omega_3 - \omega_2, 0\}].$$

Thus the total value of completing high school $E_1[\tilde{V}_2]$ includes both the direct value of a high school degree $E_1[X_2'\beta_2 + \alpha_2' E_2[\theta] + \omega_2]$ and the value of the option to attend college $E_1[\max\{X_s'(\beta_3 - \beta_2) + (\alpha_3 - \alpha_2)' E_2[\theta] + \omega_3 - \omega_2, 0\}]$. Using this notation the student graduates from high school if $E_1[\tilde{V}_2] > E_1[V_1]$, and conditional on high school graduation, he attends college if $E_2[V_3] > E_2[V_2]$.

For identification we need a location normalization (see Taber, 2000), so I set $V_2 = 0$. Then

$$E_1[\tilde{V}_2] \equiv E_1[\max\{X_3'\beta_3 + \alpha_3' E_2[\theta] + \omega_3, 0\}],$$

and I can interpret $E_1[\tilde{V}_2]$ as the value of the option to attend college.

Computation is simplified by assuming that ω_1 and ω_3 are normally distributed with mean zero and that they are independent of X and θ. Furthermore, I assume that ω_3 is independent of anything in the agent's information set during the first period. Thus the agents know the distribution of ω_3 during the first time period, but do not observe its value until the second. I simplify the notation by writing $\theta^s \equiv E^2[\theta]$. I also assume that θ^s is known to the student during the first time period. This assumption is by no means necessary, but cannot be rejected in the data and simplifies the exposition (and the computation) considerably.[8] I also attempted to relax the strict restriction on the agent's information about ω_3 by allowing the students to have some private but imperfect knowledge about ω_3 during the first period, but this restriction also cannot be rejected in the data.

8. To test this restriction I added another independent unobservable that entered the wage equation and the college decision equation, but had no influence on the high school decision. The t-statistic for that factor in the college decision was 0·45.

It is easiest to solve the model working backwards. I define d_s to be a dummy variable equal to one if the individual chooses schooling level s, and 0 otherwise. Conditional on graduating from high school, the probability of attending college takes the form

$$\Pr(d_3 = 1 \mid X, \theta^s, d_1 = 0) = \Phi(X_3'\beta_{31} + \alpha_3^{s\prime}\theta^s).$$

To solve for the probability of graduating from high school, we must first solve for $E_1[\tilde{V}_2]$. Using the assumptions above one can show that

$$E_1[\tilde{V}_2] = \sigma_3\left[\Phi\left(\frac{X_3'\beta_3 + \alpha_3^{s\prime}\theta^s}{\sigma_3}\right)\left(\frac{X_3'\beta_3 + \alpha_3^{s\prime}\theta^s}{\sigma_3}\right) + \phi\left(\frac{X_3\beta_3 + \alpha_3^{s\prime}\theta^s}{\sigma_3}\right)\right],$$

where Φ and ϕ are the standard normal cdf and pdf and σ_3 is the standard deviation of ω_3. For simplicity I redefine the parameters β_3 and α_3^s by dividing them by σ_3. I also normalize the variance of ω_1 to unity. The probability of dropping out of high school is

$$\Pr(d_1 = 1 \mid X, \theta^s) = \Phi(X_1'\beta_1 - \alpha_1^s\theta^s - E_1[\tilde{V}_2]).$$

I complete the model by assuming that log wages at time t with schooling level j take the form

$$\log(W_{tj}) = Z_{tj}'\gamma_{tj} + \delta_{tj}'\theta + \varepsilon_{tj}, \tag{5}$$

where ε_{tj} is iid $N(0, \sigma_{\varepsilon jt}^2)$ and the variance is unrestricted across time and schooling sector.

5.2. *Estimation*

Since the focus of this paper is the college/high school differential I do not estimate the parameters of the wage equation for high school dropouts. Define d_1, d_2 and d_3 to be indicators of high school dropouts, high school graduates, and college attenders respectively. In practice I assume that θ is two dimensional with the first dimension (θ^s) known during schooling and the second (θ^w) independent of information known during high school so that I can write the wage equation as

$$\log(W_{tj}) = Z_{tj}'\gamma_{tj} + \delta_{tj}^s\theta^s + \delta_{tj}^w\theta^w + \varepsilon_{tj}. \tag{6}$$

This second type of unobserved ability (θ^w) is analogous to a standard random effect in panel data models. It could result from uncertainty about job matching, about sector specific ability, or about human capital acquisition while in college.

Let $G^s(\theta^s)$ and $G^w(\theta^w)$ denote the distributions of θ^s and θ^w which are independent of $(X_1, X_2, X_3, Z_{t1}, Z_{t2}, Z_{t3})$. The likelihood for an individual is

$$\begin{aligned}\mathscr{L}(W_{ts}, d_1, d_2, d_3) = \int\int & d_1\{\Phi(X_1'\beta_1 + \alpha_1^s\theta^s - E_1[\tilde{V}_2])\} \\ & + d_2\Bigg\{(1 - \Phi(X_1'\beta_1 + \alpha_1^s\theta^s - E_1[\tilde{V}_2]))(1 - \Phi(X_3'\beta_3 + \alpha_3^s\theta^s)) \\ & \times \prod_{t=1}^{T}\phi\left(\frac{\log(W_{t2}) - Z_{t2}'\gamma_{t2} - \delta_{t2}^s\theta^s - \delta_{t2}^w\theta^w}{\sigma_{\varepsilon_{2t}}}\right)\Bigg\} \\ & + d_3\Bigg\{(1 - \Phi(X_1'\beta_1 + \alpha_1^s\theta^s - E_1[\tilde{V}_2]))\Phi(X_3'\beta_3 + \alpha_3^s\theta^s) \\ & \times \prod_{t=1}^{T}\phi\left(\frac{\log(W_{t3}) - Z_{t3}'\gamma_{t3} - \delta_{t3}^s\theta^s - \delta_{t3}^w\theta^w}{\sigma_{\varepsilon_{3t}}}\right)\Bigg\} dG^s(\theta^s)dG^w(\theta^w). \end{aligned} \tag{7}$$

The model can be estimated parametrically by specifying G^s and G^w up to a finite number of parameters or by Nonparametric Maximum Likelihood as in, for example, Heckman and Singer (1984).

In most of the results that follow I restrict θ^s and θ^w to be normal random variables. They are independent by definition since θ^w is mean zero conditional on information known during school. I evaluate the likelihood function using Gauss–Newton quadrature points. I can normalize the mean and variance of θ^s and θ^w to any values by redefining the intercepts and factor loading terms (*i.e.* θ^s_{tj}, δ^w_{tj} and a^2_j) appropriately. For computational reason it was convenient to normalize them to have mean 0·5 and variance 0·02.

5.3. *Marginal treatment effects*

A problem arises in presenting the results from this estimated model. For a given individual the gain in wages from attending college is well defined. However, the return varies with unobserved ability and thus is random across individuals in the population. I need some manner in which to summarize this random variable. The typical practice in the literature on programme evaluation is to present either the "average treatment effect", $E[\log(W_{t3}) - \log(W_{t2})]$ or the expected "treatment on the treated", $E[\log(W_{t3}) - \log(W_{t2}) | s = 3]$. The problem with both of these parameters is that if we interpret them as policy counter-factuals as in Ichimura and Taber (2000), both correspond to drastic policy changes. The first represents the difference in average earnings between a counter-factual world in which everyone in the population is forced to attend college versus a world in which everyone must receive nothing more nor less than a high school diploma. The second parameter corresponds to the difference between the average earnings for those individuals who actually go to college and the average earnings they would have received if they had stopped after high school. Thus this counter-factual corresponds to a policy in which college is eliminated. Neither of these policies resemble the type we are likely to consider.

Since I do not have a specific policy to evaluate, it is natural to condition on a group of people who are likely to be influenced by policy. With this in mind, I use the "marginal treatment effect" which is the expected earnings for an individual who is just indifferent about whether to attend college.[9] In this model, two groups of people are defined to be indifferent. The first is clear. It consists of students who completed high school and are indifferent between attending college and entering the labour force ($X'_3\beta_3 + \alpha^s_3\theta^s + \omega_3 = 0$). The second group includes individuals who are indifferent about completing high school ($X'_1\beta_1 + \alpha^s_1\theta^s + \omega_1 = E_2[\tilde{V}_2]$), but would attend college if they had completed high school.

6. ESTIMATION OF FULL SCHOOLING MODEL

6.1. *Main specification*

I first present a preliminary version of the model in Table 6. The model is restricted so that unobserved heterogeneity does not affect schooling decisions (*i.e.* $\alpha^s_1 = \alpha^s_3 = 0$), but

9. This parameter is somewhat similar to the local average treatment effect defined by Imbens and Angrist (1994). Heckman and Vytlacil (2000*a*) consider a closely related parameter they call local IV and discuss the relationship between these various parameters. Bjorkland and Moffitt (1987) also consider a version of a marginal treatment effect.

does enter the wage equation. The wage model is a random effects model with two dimensions of heterogeneity and factor loading terms that vary over time. I will not discuss the estimated parameters on the variables in Table 6 in detail, but the coefficients in the schooling equation and in the wage equation are similar to what have been found in other studies.

Consider identification of the skill prices δ_{tj}^{w} and δ_{tj}^{s} in this random effects model. Since all of the error terms are normally distributed, these parameters could not be identified without longitudinal data. Within each schooling group there are six of these factor loading parameters, and six covariances to identify them (3 within period and 3 across period). Thus these parameters are essentially just identified from the covariances between error terms in the data. However, in this version of the model θ^s and θ^w are interchangeable, so for each schooling sector, it is impossible to tell which group of skill prices is associated with θ^s and which is associated with θ^w. In this specification there are four equivalent local optima.

Now consider identification of the full selection model in which α_1^s and α_3^s are estimated as well. The main goal of this paper is to measure the pattern of selection dictated by the eight factor loading terms (α_1^s, α_3^s, δ_{tj}^{w}, δ_{tj}^{s}, for $j = 1, 2, 3$). However, in practice, the values of δ_{tj}^{w} and δ_{tj}^{s} seem to be determined by the covariance of wages in the data rather than from selection as can be seen by comparing Tables 6 and 7. Furthermore, since the factor loading term in the high school equation (α_1^s) does not affect selection into college directly, the key parameter for this analysis is α_3^s.

To present the intuition for identification of α_3^s in this model I again show the importance of exclusion restrictions but abstract from other covariates. To see how panel data can aid identification of the model consider the extreme case in which we have enough observations on each individual that we could actually identify θ^s up to scale and location. Ignoring the high school decision, individuals attend college if

$$\beta X_i + \alpha_3^s \theta_i^s + \omega_{3i} \geqq 0.$$

Since θ_i^s is only identified up to scale and location for each schooling group, we could never include it directly as a covariate into the probit equation since the scale and location will be different for high school and college graduates. However, conditioning on college attenders, after normalizing the scale and location of θ_i^s we could identify the joint distribution of $\beta X_i + \alpha_3^s \theta_i^s + \omega_{3i}$ and θ_i^s conditional on X_i since for any real τ we could identify

$$\Pr(\theta_i^s \leqq \tau \mid d_{i3} = 1, X_i) \Pr(d_{i3} = 1 \mid X_i)$$
$$= \Pr(\beta X_i + \alpha_3^s \theta_i^s + \omega_{3i} \geqq 0, \theta_i^s \leqq \tau \mid X_i).$$

The intuition for identification of α_3^s is analogous to identification in the selection model described above. If α_3^s were positive an individual with a very small value of βX_i would typically have to possess a large value of θ_i^s in order to have chosen to attend college. In that case, conditional on going to college, the distribution of θ_i^s shifts to the right as βX_i decreases. Thus, conditional on college, if θ_i^s and βX_i are negatively correlated, α_3^s will be positive. Analogous intuition holds for high school graduates.

The primary difference between identification in this case as opposed to the selection model is that the identification of θ_i^s depends not only on the level of wages but also on the trend. If δ_{jt}^{s} is rising over time, students with high values of θ_i^s represent the individuals whose earnings grew the most during the 1980s. That is a positive value of α_3^s would result if individuals who look as if they should not go to college, but attend anyway, have had

TABLE 6

Base schooling model without selection. Two independent dimensions of heterogeneity, each dimension distributed N(0·5, 0·02)†, NLSY white males (Standard errors in parentheses)

Dynamic programming selection equation		
Variable	Complete high school	Attend college
a^w	—	—
α^s	—	—
Intercept	−1·108 (0·015)	−1·995 (0·384)
Highest grade mother	0·078 (0·022)	0·055 (0·020)
Highest grade father	0·034 (0·018)	0·084 (0·015)
Number of siblings	−0·032 (0·021)	−0·062 (0·022)
Age January 1, 1978	0·058 (0·020)	0·029 (0·019)
General science score	0·017 (0·017)	0·051 (0·015)
Word knowledge score	0·019 (0·010)	0·026 (0·010)
Mathematics knowledge score	0·071 (0·017)	0·092 (0·008)
Auto and shop score	−0·008 (0·015)	−0·083 (0·010)
σ_ε	0·571 (0·307)	—

Regression equation Dependent variable: Log weekly wage		
Variable	College	High school
1982–1984:		
δ^w	3·404 (0·139)	3·342 (0·119)
δ^s	0·428 (0·248)	0·134 (0·197)
General science score	0·020 (0·011)	0·002 (0·007)
Word knowledge score	−0·024 (0·008)	0·004 (0·005)
Mathematics knowledge score	0·017 (0·007)	0·008 (0·005)
Auto and shop score	0·010 (0·007)	0·011 (0·005)
Highest grade × college	0·091 (0·024)	—
1985–1987:		
δ^w	2·366 (0·148)	2·617 (0·123)
δ^s	2·059 (0·155)	1·747 (0·160)
General science score	0·010 (0·007)	−0·008 (0·006)
Word knowledge score	−0·012 (0·006)	0·005 (0·004)
Mathematics knowledge score	0·020 (0·004)	0·010 (0·004)
Auto and shop score	0·001 (0·004)	0·012 (0·005)
Highest grade × college	0·091 (0·013)	—
1988–1990:		
δ^w	1·317 (0·170)	2·034 (0·160)
δ^s	3·433 (0·098)	2·809 (0·128)
General science score	0·004 (0·007)	−0·016 (0·006)
Word knowledge score	−0·014 (0·006)	0·009 (0·004)
Mathematics knowledge score	0·017 (0·004)	0·013 (0·004)
Auto and shop score	0·005 (0·004)	0·012 (0·005)
Highest grade × college	0·113 (0·011)	—
Experience	0·120 (0·008)	0·120 (0·008)
Experience squared	−0·446 (0·047)	−0·446 (0·047)
Local unemployment rate	−0·032 (0·005)	−0·032 (0·005)
Age January 1, 1978	0·004 (0·007)	0·004 (0·007)

	Average treatment effect	Marginal treatment effect
1982–1984	0·332 (0·060)	0·317 (0·060)
1985–1987	0·394 (0·041)	0·383 (0·041)
1988–1990	0·472 (0·038)	0·457 (0·038)
Log likelihood	−17,729·549	
Number of individuals	1836	

† Likelihood is calculated using 7 Gauss–Newton quadrature points to approximate normal distribution.

TABLE 7

Results for full schooling model, two independent dimensions of heterogeneity. Each dimension distributed N(0·5, 0·02)†, NLSY white males
(Standard errors in parentheses)

Dynamic programming selection equation		
Variable	Complete high school	Attend college
a^w	—	—
α^s	1·481 (1·229)	1·468 (0·637)
Intercept	−1·884 (0·015)	−2·805 (0·543)
Highest grade mother	0·080 (0·023)	0·054 (0·020)
Highest grade father	0·032 (0·019)	0·084 (0·015)
Number of siblings	−0·031 (0·022)	−0·063 (0·022)
Age January 1, 1978	0·060 (0·021)	0·033 (0·019)
General science score	0·016 (0·017)	0·053 (0·015)
Word knowledge score	0·020 (0·010)	0·028 (0·011)
Mathematics knowledge score	0·071 (0·018)	0·096 (0·009)
Auto and shop score	−0·007 (0·015)	−0·085 (0·011)
σ_ε	0·681 (0·350)	—
Regression equation Dependent variable: log weekly wage		
Variable	College	High school
1982–1984:		
δ^w	3·381 (0·155)	3·333 (0·119)
δ^s	0·171 (0·259)	0·161 (0·198)
General science score	0·007 (0·011)	0·001 (0·008)
Word knowledge score	−0·015 (0·008)	0·003 (0·005)
Mathematics knowledge score	0·016 (0·007)	0·007 (0·005)
Auto and shop score	0·018 (0·006)	0·012 (0·005)
Highest grade × college	0·115 (0·024)	—
1985–1987:		
δ^w	2·452 (0·133)	2·592 (0·122)
δ^s	1·855 (0·155)	1·769 (0·160)
General science score	0·001 (0·007)	−0·008 (0·006)
Word knowledge score	−0·004 (0·005)	0·005 (0·004)
Mathematics knowledge score	0·023 (0·004)	0·013 (0·004)
Auto and shop score	0·004 (0·004)	0·012 (0·005)
Highest grade × college	0·111 (0·013)	—
1988–1990:		
δ^w	1·508 (0·166)	1.993 (0·158)
δ^s	3·330 (0·104)	2·878 (0·131)
General science score	−0·002 (0·007)	−0·015 (0·006)
Word knowledge score	−0·006 (0·006)	0·010 (0·004)
Mathematics knowledge score	0·023 (0·004)	0·018 (0·005)
Auto and shop score	0·006 (0·004)	0·011 (0·005)
Highest grade × college	0·132 (0·011)	—
Experience	0·120 (0·008)	0·120 (0·008)
Experience squared	−0·445 (0·048)	−0·445 (0·048)
Local unemployment rate	−0·030 (0·005)	−0·030 (0·005)
Age January 1, 1978	0·001 (0·007)	0·001 (0·007)
	Average treatment effect	Marginal treatment effect
1982–1984	0·346 (0·060)	0·337 (0·060)
1985–1987	0·337 (0·043)	0·331 (0·042)
1988–1990	0·360 (0·052)	0·349 (0·051)
Log likelihood	−17,725·932	
Number of individuals	1836	

† Likelihood is calculated using 7 Gauss–Newton quadrature points to approximate normal distribution.

large wage gains during the 1980s while individuals who look as if they should go to college during, but do not, have seen smaller gains.

As mentioned above there are four different local optima associated with each configuration of factor loading terms in the random effects model. I typically find that for each schooling group there is one factor whose skill price is increasing over time and one that is falling over time. In this subsection, I restrict the model so that the dimension with an increasing factor is associated with θ^s. One can interpret this as a test of the hypothesis mentioned in the last paragraph. I relax this restriction and explore the alternatives in Section 6.4 below.

The estimation results for this version of the full structural model appear in Table 7. Since many of the covariates are interacted, the raw numbers are hard to interpret. The table can be read most clearly by looking at the summary statistics at the bottom of the second page. In terms of both the marginal treatment effect and the average treatment effect, controlling for unobserved ability eliminates the increase in the college premium. As opposed to the two-step method, I find strong evidence of an increase in the payoff to unobserved ability as can be seen either from constructing the t-statistic on α_3^s or by looking at the difference in the likelihood between this model and that on Table 6.

A curious aspect of the results is that the levels of the marginal treatment effect and the average treatment effect are quite similar. To gain a further understanding I simulated some other treatment effects and found them also to be similar to the average and marginal treatment effects. These similarities occur both because (a) the magnitude of heterogeneity in returns is fairly small[10] and (b) since about 50% of individuals attend college, the marginal college entrant has close to median unobserved ability (and with normal heterogeneity the median is the same as the mean).

6.2. *Selection model*

To test the sensitivity of the model to the dynamic programming specification, I estimate an intermediate model in Table 8. I present results from a probit selection model (conditioning on high school graduation), but estimated via full maximum likelihood with the two dimensional ability vector. This method produces treatment effects that are virtually identical to the dynamic programming specification. As in the full schooling model there is strong evidence of an increasing payoff to unobserved ability. This table can also be used to test whether I have found evidence of a dynamic programming element to the schooling selection process. Since this model is a special case of the dynamic programming model,[11] I can test the specification using a likelihood ratio test. When I perform this test, I reject the standard model in favour of the full schooling model at the 10% level, but not at the 5% level.

6.3. *Relaxing the normality assumption*

Since these results may be sensitive to the functional forms used, I attempted to relax the distributional assumptions on θ^s and θ^w. I first attempted to estimate the model using

10. Note that the fact that there is not a lot of heterogeneity in the returns to college does not necessarily mean that unobserved heterogeneity is not important. If the coefficient on unobserved heterogeneity (δ_{tj}^s) were the same in the college sector as the high school sector, then the "returns to college" would not vary with θ^s, but there still may be substantial bias as θ^s may be strongly correlated with schooling and wages.

11. The dynamic programming model reduces to the selection model if the coefficients of α_1^* and σ_3 in the complete high school equation are both equal to zero.

TABLE 8

Results for selection model. Two independent dimensions of heterogeneity, each dimension distributed $N(0{\cdot}5, 0{\cdot}02)$†, *NLSY white males*
(*Standard errors in parentheses*)

Selection equation		
Variable	Attend college	
α^w	—	
α^s	1·648 (0·632)	
Intercept	−2·843 (0·532)	
Highest grade mother	0·054 (0·020)	
Highest grade father	0·085 (0·015)	
Number of siblings	−0·065 (0·023)	
Age January 1, 1978	0·031 (0·019)	
General science score	0·054 (0·015)	
Word knowledge score	0·028 (0·011)	
Mathematics knowledge score	0·095 (0·009)	
Auto and shop score	−0·086 (0·011)	
Regression equation Dependent variable: log weekly wage		
Variable	College	High school
1982–1984:		
δ^w	3·389 (0·156)	3·337 (0·118)
δ^s	0·153 (0·258)	0·139 (0·190)
General science score	0·006 (0·011)	0·002 (0·007)
Word knowledge score	−0·015 (0·008)	0·003 (0·005)
Mathematics knowledge score	0·017 (0·007)	0·007 (0·005)
Auto and shop score	0·018 (0·006)	0·012 (0·005)
Highest grade × college	0·117 (0·024)	—
1985–1987:		
δ^w	2·467 (0·130)	2·595 (0·116)
δ^s	1·834 (0·155)	1·749 (0·152)
General science score	0·001 (0·007)	−0·007 (0·005)
Word knowledge score	−0·004 (0·005)	0·004 (0·004)
Mathematics knowledge score	0·022 (0·004)	0·012 (0·004)
Auto and shop score	0·004 (0·004)	0·012 (0·005)
Highest grade × college	0·112 (0·013)	—
1988–1990:		
δ^w	1·531 (0·164)	1·997 (0·153)
δ^s	3·315 (0·104)	2·865 (0·122)
General science score	−0·003 (0·007)	−0·014 (0·006)
Word knowledge score	−0·006 (0·006)	0·009 (0·004)
Mathematics knowledge score	0·022 (0·004)	0·016 (0·004)
Auto and shop score	0·005 (0·004)	0·011 (0·005)
Highest grade × college	0·133 (0·011)	—
Experience	0·121 (0·008)	0·121 (0·008)
Experience squared	−0·457 (0·048)	−0·457 (0·048)
Local unemployment rate	−0·030 (0·005)	−0·030 (0·005)
Age January 1, 1978	−0·001 (0·007)	−0·001 (0·007)
	Average treatment effect	Marginal treatment effect
1982–1984	0·345 (0·058)	0·341 (0·059)
1985–1987	0·334 (0·042)	0·326 (0·042)
1988–1990	0·339 (0·051)	0·334 (0·050)
Log likelihood‡	−17,728·327	
Number of individuals	1527	

† Likelihood is calculated using 7 Gaus–Newton quadrature points to approximate normal distribution.
‡ The likelihood also includes a probit for high school selection so the likelihood can be directly compared with the other specifications.

nonparametric maximum likelihood as in Heckman and Singer (1984) but this did not work for computational reasons. Instead, I test the robustness of the model by allowing the distribution of the heterogeneity to take several other forms. This procedure is less than ideal. A nice feature of the model is that when θ^s, θ^w and ε_{ts} are normally distributed, the factor loading terms δ^s_{tj} and δ^w_{tj} are identified from covariances in the panel data. When θ^s is not normal this is no longer the case and panel data is not required for identification of the factor loading terms. When ε_{ts} is restricted to be normal, but the marginal distribution of wages are not, δ^s_{tj} can be identified from departures from normality in the marginal distribution of wages. Similarly, when the error terms of the selection model is restricted to be normal, identification of α^s can also come from normality in the selection equation. This makes the estimates more difficult to interpret.[12]

TABLE 9

Results for full schooling model. Two independent dimensions of heterogeneity, alternative heterogeneity distributions, NLSY white males
(Standard errors in parentheses)

	Dynamic programming selection equation					
	Mixture of normals 1†		Mixture of normals 2†		Uniform‡	
Variable	Complete high school	Attend college	Complete high school	Attend college	Complete high school	Attend college
α^w	—	—	—	—	—	—
α^s	−0·049 (1·424)	2·122 (0·665)	0·100 (1·655)	7·393 (1·427)	2·948 (0·654)	−3·271 (0·396)
	Regression equation					
Variable	College	High school	College	High school	College	High school
1982–1984:						
δ^w	4·012 (0·200)	3·999 (0·212)	3·574 (0·209)	3·667 (0·154)	1·195 (0·059)	1·191 (0·041)
δ^s	1·586 (0·346)	1·730 (0·302)	−1·461 (0·342)	1·031 (0·353)	−0·197 (0·094)	0·692 (0·078)
1985–1987:						
δ^w	2·303 (0·234)	2·494 (0·254)	3·326 (0·140)	2·546 (0·212)	1·161 (0·052)	0·909 (0·053)
δ^s	3·228 (0·187)	3·187 (0·197)	1·071 (0·288)	2·782 (0·260)	0·566 (0·075)	1·414 (0·070)
1988–1990:						
δ^w	0·514 (0·248)	1·153 (0·316)	2·936 (0·229)	1·612 (0·305)	1·128 (0·069)	0·662 (0·051)
δ^s	4·476 (0·101)	4·298 (0·137)	3·140 (0·259)	4·019 (0·194)	1·137 (0·077)	1·813 (0·077)
	Average treatment effect	Marginal treatment effect	Average treatment effect	Marginal treatment effect	Average treatment effect	Marginal treatment effect
1982–1984	0·269 (0·067)	0·264 (0·066)	0·331 (0·073)	0·297 (0·079)	0·501 (0·065)	0·519 (0·066)
1985–1987	0·260 (0·048)	0·261 (0·046)	0·126 (0·055)	0·107 (0·063)	0·887 (0·051)	0·909 (0·054)
1988–1990	0·288 (0·050)	0·287 (0·050)	−0·006 (0·071)	−0·018 (0·079)	1·175 (0·059)	1·190 (0·063)
Likelihood	−17,700.106		−17,731·122		−17,897·273	
Individ.	1836		1836		1836	

† Both mixtures use 8 Gauss–Newton quadrature points to approximate normal distribution. The first has expected value zero for both distributions, but the second has double the variance. The second mixture has the same variance for each, but the expected values are one standard deviation apart.
‡ The uniform is approximated with 5 equally spaced points with equal probability.

The estimates allowing for alternative forms for the distribution of the ability terms are presented in Table 9. The first specification is a mixture of normals each with equal probability and zero expected value, with one having twice the variance of the first. These results are presented in the first two columns. They yield treatment effects quite similar to Table 7 showing strong evidence of an increase in the return to ability and only a slight

12. This may also potentially be part of the problem with the semiparametric specification above. The identification of distribution of the heterogeneity terms may come from the marginal distribution of wages rather than the covariance of wages as one might like.

increase in the return to schooling. The likelihood of this model is higher than the base model of Table 7. The next specification allows for a mixture of normals with equal probability and the same variance, but the expected values were a standard deviation apart. This specification was different than any of the others I have tried both in that there were multiple local optima in the model without selection, and that the factor loading coefficient on the ability term switches sign from the first to the second period. This model finds very strong evidence of an increasing payoff to ability; so strong that the trend in the return to schooling is actually reversed. The likelihood of this model is lower than the base model. The third specification allows for essentially a uniform distribution for the unobserved heterogeneity. This model actually shows a declining return to ability, but the likelihood of this model is much worse than the standard model.

6.4. *Relaxing the restrictions*

In this final subsection I relax the restrictions in the main specification along a number of dimensions. Table 6 presents results from a two factor model in which both factors are treated symmetrically. In implementing the primary specification in Table 7, I allow the second factor to influence schooling choices but not the first. In results presented in Table 10, I employ an alternative restriction by allowing the first factor to influence schooling rather than the second. Again, we see that most (although not all) of the increase in the schooling premium is eliminated when we account for selection. However, it differs from the main specification in that the level of the selection bias switches sign; the return become higher than the OLS returns rather than lower (as in the IV results above). The

TABLE 10

Results for alternative restriction: skill price falling over time, coefficients on skills (Standard errors in parentheses)

Dynamic programming selection equation		
Variable	Complete high school	Attend college
α^w	—	—
α^s	−2·227 (1·127)	−1·884 (0·579)
Regression equation		
Variable	College	High school
1982–1984:		
δ^s	3·489 (0·148)	3·458 (0·123)
δ^w	0·658 (0·211)	−0·006 (0·173)
1985–1987:		
δ^s	2·439 (0·136)	2·831 (0·111)
δ^w	2·195 (0·156)	1·610 (0·142)
1988–1990:		
δ^s	1·320 (0·152)	2·264 (0·148)
δ^w	3·524 (0·090)	2·682 (0·115)
	Average treatment	Marginal treatment
1982–1984	0·496 (0·069)	0·484 (0·069)
1985–1987	0·522 (0·049)	0·515 (0·048)
1988–1990	0·557 (0·042)	0·548 (0·042)
Log likelihood	−17,725·699	

log likelihood of this specification is very similar to the log likelihood for the main specification. Following the identification argument above, this result is consistent with the main specification above if θ_i^s is primarily identified by the trend in wages rather than the levels. Both results suggest that the individuals who see their wages rise most rapidly are more likely to attend college.

Relaxing the model along a different dimension gives results that are not as robust and more difficult to interpret. Returning once again to the model without selection, there is no relationship between the increasing factor in the college wage equation and the increasing factor in the high school wage equation. There is no *a priori* reason to restrict them to have the same coefficient in the schooling selection equations. I relax this restriction in the model presented in Table 11. This model has three factors, one that does not influence schooling, one that enters the college wage equation but not the high school wage equation, and another that enters the high school wage equation but not the college wage equation. This yields two local optima which are essentially indistinguishable from

TABLE 11

Results for alternative restriction: allowing differential selection for high school and college skills. Three independent dimensions of heterogeneity, each dimension distributed N(0·5, 0·02)†, NLSY white males (Standard errors in parentheses)

Dynamic programming selection equation				
	Local Opt. 1		Local Opt. 2	
Variable	Complete high school	Attend college	Complete high school	Attend college
α^w	—	—	—	—
α^{sH}	1·246 (1·523)	1·662 (1·040)	1·210 (1·415)	1·913 (1·110)
α^{sC}	3·391 (3·508)	1·151 (1·088)	1·726 (3·600)	−2·900 (1·235)
Regression equation: Dependent variable: log weekly wage				
Variable	College	High school	College	High school
1982–1984:				
δ^w	3·381 (0·155)	3·334 (0·119)	3·512 (0·151)	3·336 (0·119)
δ^{sH}	—	0·153 (0·197)	—	0·147 (1·195)
δ^{sC}	0·166 (0·259)	—	0·640 (0·233)	—
1985–1987:				
δ^w	2·453 (0·135)	2·595 (0·121)	2·388 (0·131)	2·597 (0·120)
δ^{sH}	—	1·761 (0·158)	—	1·759 (0·156)
δ^{sC}	1·861 (0·157)	—	2·237 (0·162)	—
1988–1990:				
δ^w	1·510 (0·170)	1·995 (0·157)	1·302 (0·143)	2·000 (0·155)
δ^{sH}	—	2·874 (0·129)	—	2·880 (0·128)
δ^{sC}	3·343 (0·113)	—	3·586 (0·102)	—
	Average treatment effect	Marginal treatment effect	Average treatment effect	Marginal treatment effect
1982–1984	0·347 (0·060)	0·337 (0·060)	0·387 (0·063)	0·367 (0·062)
1985–1987	0·333 (0·043)	0·327 (0·043)	0·459 (0·044)	0·432 (0·044)
1988–1990	0·351 (0·052)	0·340 (0·051)	0·557 (0·050)	0·516 (0·051)
Log likelihood	−17,725·157		−17,725·934	
Number of individuals	1836		1836	

† Likelihood is calculated using 7 Gauss–Newton quadrature points to approximate normal distribution.

each other in that changing the specification of the model will change which optimum is lower.[13] The first local optimum is very similar to the primary specification in Table 7. We see a positive coefficient on both "ability" terms in the schooling decisions that eliminates the increase in the college premium. The second local optimum is different. In this case the high school factor enters positively into the college decision as before, but the college factor enters negatively. The average return to college actually rises faster than in OLS. In both optima we see evidence of an increasing payoff to ability in the high school sector. However, for the college sector we see this evidence in one local optimum, but get the opposite result in the other.

It is not clear exactly how to proceed in this case. Econometric theory is essentially silent. One can obtain a consistent estimate of the model by choosing the optimum with the highest likelihood. However, this may be very misleading as one cannot tell which is the "true" optima. This problem does not appear to be common. Other examples include Ham and Altonji (1990); Hotz and Miller (1993); Ham, Svejnar and Terrell (1998) who also recognize the problem but do not have a clear solution to it.

In the final specification presented in Table 12, I return to the two factor model, but allow both factors to influence schooling. I again obtain two local optima in which the likelihood is close. The second local optimum gives results very similar to the main results in that the second factor is positive as in Table 7, and the first factor is negative as in Table 9. Both lead to a fall in the return to college, so once again it is essentially flat over time. However, the first local optimum results in a negative coefficient on unobserved ability in the attend college equation for both cases, and the college premium increases faster than in OLS.[14]

7. CONCLUSION

In both the instrumental variable model and the Heckman two step specification I find no evidence of an increasing payoff in the causal effect of schooling, although the standard errors are very large. The results from the structural model yield more precise estimates. They typically also find no evidence of an increasing college premium. However, interpreting the results from the structural model is complicated by multiple optima. In general, when I allow one factor to influence schooling and wages I find that selection on this factor can explain much of the increase in the measured return to schooling over time. Allowing for two factors yields two local optima, one in which both effects go in the same direction toward reducing the increase in the pattern over time, and another in which the effects go in opposite directions.

Even if the causal effect of schooling disappears, it is difficult to make policy predictions based on these results since the intrinsic skills with increasing value are unobservable. However, we have some information about them. They are not well proxied by the ASVAB scores and they do not appear to be learned in college. Furthermore, the skills are either learned before leaving school or are highly correlated with information that the student possesses while in school. One possibility is that they are related to motivation or ability to learn. Hopefully, future work will shed light on the question.

13. For example changing the way in which cohort effects enter the model does not lead to big differences in each of the local optima, but does change which optimum has the higher likelihood.

14. In the basic model one also obtain two local optima, but the difference between the two optima is much larger (2·4 log-likelihood points). The only case in which these log-likelihood are close occurs when cohort effects are completely unrestricted. The other local optimum gives unreasonable imprecisely estimated cohort effects in the wage equation, so the smoothing seems reasonable.

TABLE 12

Results for alternative restriction: both factors influence schooling. Two independent dimensions of heterogeneity. Each dimension distributed N(0·5, 0·02)† NLSY white males (Standard errors in parentheses)

Dynamic programming selection equation				
	Local Opt. 1		Local Opt. 2	
Variable	Complete high school	Attend college	Complete high school	Attend college
α^{s_1}	−2·695 (1·549)	−2·027 (0·618)	−2·352 (1·596)	−0·781 (0·758)
α^{s_2}	3·382 (1·549)	−1·910 (0·618)	2·787 (1·596)	1·962 (0·758)
Regression equation: Dependent variable: log weekly wage				
Variable	College	High school	College	High school
1982–1984:				
δ^{s_1}	3·506 (0·148)	3·489 (0·123)	3·386 (0·154)	3·382 (0·122)
δ^{s_2}	0·687 (0·203)	−0·024 (0·176)	0·184 (0·258)	0·130 (0·193)
1985–1987:				
δ^{s_1}	2·409 (0·132)	2·929 (0·103)	2·452 (0·131)	2·606 (0·120)
δ^{s_2}	2·248 (0·162)	1·607 (0·139)	1·844 (0·151)	1·746 (0·159)
1988–1990:				
δ^{s_1}	1·279 (0·148)	2·385 (0·145)	1·495 (0·165)	1·955 (0·159)
δ^{s_2}	3·570 (0·092)	2·799 (0·121)	3·322 (0·105)	2·898 (0·132)
	Average treatment effect	Marginal treatment effect	Average treatment effect	Marginal treatment effect
1982–1984	0·511 (0·071)	0·501 (0·070)	0·347 (0·060)	0·337 (0·060)
1985–1987	0·620 (0·047)	0·617 (0·047)	0·333 (0·043)	0·327 (0·043)
1988–1990	0·718 (0·049)	0·713 (0·048)	0·351 (0·052)	0·340 (0·051)
Log likelihood	−17,722·224		−17,723·259	
Number of individuals	1836		1836	

† Likelihood is calculated using 7 Gauss–Newton quadrature points to approximate normal distribution.

On its own, this evidence may not be convincing that there has been an increase in the return to ability over time that explains the increasing schooling premium. However, it suggests that this could be a very important part of the story. While the evidence of an increase in the payoff to these unobserved skills may be weak, the evidence that the causal effect of schooling increased is even weaker.

Acknowledgements. This paper constituted chapter three of my Ph.D. dissertation at the University of Chicago. I would like to thank Joe Altonji, Orazio Attanasio, Bo Honoré, Joe Hotz, Hide Ichimura, Bruce Meyer, Paul Oyer, Sherwin Rosen, Brian Tyler and several referees for helpful comments, and to Steve Cameron for comments and for providing me with some of the code. I am especially grateful to James Heckman for many useful conversations. I also thank workshop participants at a number of institutions. I gratefully acknowledge the financial support of the Searle Foundation and the Alfred P. Sloan Foundation. All remaining errors are my own.

REFERENCES

ACEMOGLU, D. (1998), "Why do New Technologies Complement Skills? Directed Technical Change and Wage Inequality", *Quarterly Journal of Economics*, **13**, 1055–1090.

ALTONJI, J. (1993), "The Demand for and Return to Education when Education Outcomes are Uncertain", *Journal of Labor Economics*, **11**, 48–83.

ALTONJI, J., ELDER, T. and TABER, C. (2000), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools" (Unpublished manuscript).
AMEMIYA, T. (1985) *Advanced Econometrics* (Cambridge: Harvard University Press).
AUTOR, D. and KATZ, L. (2000), "Changes in the Wage Structure and Earnings Inequality", in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* (Amsterdam: North Holland).
AUTOR, D., KATZ, L. and KRUEGER, A. (1998), "Computing Inequality: Have Computers Changed the Labor Market?", *Quarterly Journal of Economics*, **113**, 1169–1214.
BARTEL, A. and SICHERMAN, N. (1999), "Technological Change and Wages: An Interindustry Analysis", *Journal of Political Economy*, **107**, 285–325.
BEAUDRY, P. and GREEN, D. (1998), "What is Driving U.S. and Canadian Wages: Exogenous Technical Change or Endogenous Choice of Technique?" (NBER Working Paper No. 6853).
BISHOP, J. (1991), "Achievement, Test Scores, and Relative Wages", in M. Kosters (ed.), *Workers and their Wages: Changing Patterns in the United States* (Washington: AEI Press).
BJORKLAND, A. and MOFFITT, R. (1987), "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models", *Review of Economics and Statistics*, **69**, 42–49.
BLACKBURN, M. and NEUMARK, D. (1993), "Omitted-Ability Bias and the Increase in the Return to Schooling", *Journal of Labor Economics*, **11**, 521–544.
BUCHINSKY, M. and LESLIE, P. (1999), "Educational Attainment and the Changing U.S. Wage Structure: Dynamic Implications without Rational Expectations" (Unpublished manuscript).
CAMERON, S. and HECKMAN, J. (1993), "The Nonequivalence of High School Equivalents", *Journal of Labor Economics*, **11**, 1–47.
CAMERON, S. and HECKMAN, J. (1998*a*), "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, **106**, 262–333.
CAMERON, S. and HECKMAN, J. (1998*b*), "The Dynamics of Education Attainment for Blacks, Whites, and Hispanics" (Columbia University manuscript).
CAMERON, S. and TABER, C. (2000), "Borrowing Constraints and the Returns to Schooling" (Unpublished manuscript).
CARD, D. (2000), "The Causal Effect of Education on Earnings", in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* (Amsterdam: North Holland).
CARD, D. and LEMIEUX, T. (1993), "Wage Dispersion, Returns to Skill, and Black–White Wage Differentials" (NBER Working Paper No. 4365).
CAWLEY, J., HECKMAN, J., LOCHNER, L. and VYTLACIL, E. (2000), "Understanding the Role of Cognitive Ability in Accounting for the Recent Rise in the Economic Return to Education", in K. Arrow, S. Bowles and S. Durlauf (eds.), *Meritocracy and Economic Inequality* (Princeton: Princeton University Press).
CHAY, K. and LEE, D. (1999), "Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black–White Differentials" (Unpublished manuscript).
COMAY, Y., MELNIK, A. and POLLACHECK, M. (1973), "The Option Value of Education and the Optimal Path for Investment in Human Capital", *International Economic Review*, **14**, 421–435.
GRILICHES, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, **45**, 1–22.
GRILICHES, Z. (1979), "Sibling Models and Data in Economics: Beginnings of a Survey", *Journal of Political Economy*, **87**, s37–s64.
HECKMAN, J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, **47**, 153–162.
HECKMAN, J. (1990), "Varieties of Selection Bias", *American Economic Review*, **80**, 313–318.
HECKMAN, J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator", *Journal of Human Resources*. **32**, 1–40.
HECKMAN, J. and SINGER, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Economic Models for Duration Data", *Econometrica*, **52**, 271–320.
HECKMAN, J. and VYTLACIL, E. (2000*a*), "Local Instrumental Variables", in C. Hsiao, K. Morimune and J. Powell (eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya* (Cambridge: Cambridge University Press).
HECKMAN, J. and VYTLACIL, E. (2000*b*), "Cognitive Ability and the Rising Return to Education", *Review of Economics and Statistics* (forthcoming).
HERRNSTEIN, R. and MURRAY, C. (1994) *The Bell Curve: Intelligence and Class Structure in American Life* (New York: Free Press).
HOTZ, J. and MILLER, R. (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, **60**, 497–530.
ICHIMURA, H. and TABER, C. (2000), "Estimation of Policy Effects Under Limited Support Conditions" (Unpublished manuscript).
IMBENS, G. and ANGRIST, J. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, **62**, 467–475.
JUHN, C., MURPHY, K. and PIERCE, B. (1993), "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, **101**, 410–442.
KATZ, L. and MURPHY, K. (1992), "Changes in Relative Wages, 1963–1987: Supply and Demand Factors", *Quarterly Journal of Economics*, **107**, 35–78.

KEANE, M. and WOLPIN, K. (1997), "The Career Decisions of Young Men", *Journal of Political Economy*, **105**, 473–522.

KRUEGER, A. (1993), "How Have Computers Changed the Wage Structure? Evidence from Micro Data", *Quarterly Journal of Economics*, **108**, 33–60.

LEVY, F. and MURNANE, R. (1992), "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations", *Journal of Economic Literature*, **30**, 1331–1381.

MURNANE, R., WILLETT, J. and LEVY, F. (1995), "The Growing Importance of Cognitive Skills in Wage Determination", *Review of Economics and Statistics*, **77**, 251–266.

NEAL, D. and JOHNSON, W. (1996), "The Role of Premarket Factors in Black–White Wage Differences", *Journal of Political Economy*, **104**, 869–895.

TABER, C. (1998), "The Rising College Premium in the Eighties: Return to College or Return to Ability" (Unpublished manuscript).

TABER, C. (2000), "Semiparametric Identification and Heterogeneity in Dynamic Programming Discrete Choice Models", *Journal of Econometrics*, **96**, 201–229.

WILLIS, R. (1986), "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions", in O. Ashenfelter and F. Layard (eds.), *Handbook of Labor Economics* (Amsterdam: North Holland).

WILLIS, R. and ROSEN, S. (1979), "Education and Self-Selection", *Journal of Political Economy*, **87**, 507–536.