OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 65, 1 (2003) 0305-9049

Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability–earnings Relationships[1]

JUSTIN L. TOBIAS

Department of Economics, University of California-Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA. E-mail: jtobias@uci.edu

Abstract

In this paper I explore the relationship between ability and log waves using flexible estimation techniques. I find evidence of nonlinearities in these relationships that vary across levels of schooling, and argue that ability-sorting into higher education creates problems for accurately identifying the return to schooling over the entire ability support. Over an ability support that is "common" to those with and without a college education, I find that the college log wage premium is incerasing for the *more* able, and this premium grew during the period 1984–1994 for individuals at all points in the ability distribution. The growth of this wage premium over time also appears to have followed a "smoother" linear path for high-ability individuals than individuals of lower ability.

I. Introduction

There has been a tremendous amount of work in the economics literature conducted on the estimation of the private returns to education. Stemming from the innovative work of Mincer (1974), the most basic of these 'schooling' models propose that the log wage is linearly related to years of schooling completed, as well as a quadratic term in labour market experience.

[1]The author is Assistant Professor of Economics at the University of California-Irvine. I would like to thank Duncan Black, David Brownstone, John DiNardo, Kaku Furuya, Dale Poirier, Sarah Senesky, Adonis Yatchew, an anonymous referee and the Editor for their very helpful comments and suggestions. All errors are, of course my own.

© Blackwell Publishing Ltd, 2003. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

The estimated coefficient on the schooling variable in these (typically) cross-sectional regressions is commonly interpreted as the economic return to education.

In subsequent work, the standard Mincerian wage equation has seen a number of generalizations.[2] Some of these generalizations include the use of procedures to control for self-selection into higher education (e.g. Willis and Rosen, 1979), or to account for potential nonlinearities upon degree completion (e.g. Hungerford and Solon, 1987; Belman and Heywood, 1991; Jaeger and Page, 1996). Most importantly for our purposes, numerous studies have described the need to correct or control for other confounding factors such as individual-level 'ability'. Cognitive ability is a characteristic that is strongly correlated with the quantity of schooling attained and is also likely to be correlated with observed earnings. As a result, failure to account for ability in these regression relationships (holding all else constant) should result in an upward bias in reported ordinary least square (OLS) estimates of the returns to education.[3]

A number of clever approaches have been advanced to deal with the problem of ability bias in education studies. 'Twins studies', for example, (e.g. Ashenfelter and Krueger, 1994; Altonji and Dunn, 1996; Behrman, Rosenzweig and Taubman, 1996; Ashenfelter and Rouse, 1998; Rouse, 1999) use data on identical twins or siblings to difference out unobserved ability and family components to obtain estimates of the returns to education. Other studies (e.g. Weisbrod and Karpoff, 1968; Ashenfelter and Mooney, 1968; Hansen, Weisbrod and Scanlon, 1970; Griliches and Mason, 1972; Griliches, 1977; Blackburn and Neumark, 1992, 1993; Herrnstein and Murray, 1994; Murnane, Willett and Levy, 1995; Grogger and Eide, 1995; Cawley *et al.*, 1997; Cawley, Heckman and Vytlacil, 1999, 2000; Heckman and Vytlacil, 2001) simply find detailed test score information and include these test score variables in their wage regressions to obtain estimates of the return to education after directly controlling for measured cognitive ability. Instrumental variable studies and 'natural experiments' (e.g. Angrist and Krueger, 1991; Card, 1995; Harmon and Walker, 1995, 1999; Conneely and Uusitalo, 1997; Ichino and Winter-Ebmer, 1999; Meghir and Palme, 1999) seek to find exogenous sources of information that are correlated with the quantity of schooling but are uncorrelated with ability or other unobservables affecting earnings, and then use this information to obtain consistent estimates of the schooling coefficient.[4] Finally, the general use of selection models in this context

[2]For detailed discussion, see Heckman, Lochner and Todd (2001).

[3]See Blackburn and Neumark (1995) for additional discussion.

[4]Perhaps the most celebrated of these IVs are Angrist and Krueger's (1991) quarter of birth, and a proximity to college variable as used by Card (1995) and Conneely and Uusitalo (1997).

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(e.g. Heckman, 1976; Heckman, Tobias and Vytlacil, 2001), stems from the need to control for unobservables like ability that affect both the quantity of schooling attained and the observed log wage.

To be sure, the problem of ability bias is just one of many problems that potentially arise in the estimation of private returns to schooling. In thorough reviews of the literature,[5] Card (1999, 2001) discusses issues of measurement error in reported education, the possibility of individual-level heterogeneity in costs of and returns to schooling, and issues that arise with the use of IV methods when returns to education vary across individuals. Card (1999) concludes his article by mentioning a few issues that are as yet unresolved, and offers several potential directions for further research including: (1) estimating the *social* returns to education,[6] (2) disentangling productivity shifts from other explanations like job access or signalling in explaining why returns to education exist, (3) finding observables that explain variation in returns to education across individuals,[7] and (4) characterizing variation in returns to education over time.[8]

In this paper we do not address the first two issues suggested by Card, but contribute to this literature by focusing on the third suggestion and briefly commenting on the fourth. The primary purpose of this paper is to explore the role of individual-level ability in explaining variation in returns to schooling, where returns are measured by the US college–high school wage gap. We explore these relationships by taking data from the National Longitudinal Survey of Youth (NLSY). The NLSY is a rich US panel data set containing information on the labour market experiences of young men and women aged 14–22 in 1979, and we obtain data through 1994. Importantly, for our purposes, the NLSY provides a rich set of test scores from the Armed Services Vocational Aptitude Battery (ASVAB) for the vast majority of the sample, thus enabling us to investigate the role of ability in explaining the college wage premium over the sample period.

To focus ideas on modelling the evolution of the college wage gap, we will partition the data into two 'states' according to those with and without a college education. Within each of these states, we will introduce a regression equation and permit our ability variable to enter that regression equation very flexibly. We will define our college wage premium as the mean of the

[5]For other recent literature reviews, see Harmon, Oosterbeek and Walker (2002) and Sianesi and Van Reenen (2002).

[6]See Moretti (2002) for some progress on this issue.

[7]Some studies, such as Card and Krueger (1992) and Tobias and Li (2002), have examined if measures of the quality of schooling can explain some of the observed heterogeneity in returns to education, while Ashenfelter and Rouse (1998) analyse the effect of family background characteristics on schooling returns.

[8]See, for example, Grogger and Eide (1995), Murnane, Willett and Levy (1995), Heckman and Vytlacil (2001) and Taber (2001) for investigation of this issue.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

difference between the 'college' and 'no-college' log wages and will examine how this premium varies over the ability support. If this difference increases with ability, then we will find that returns to education – as measured by the average difference between the college and no-college log wages – are indeed concentrated among the more able.

The models we employ are quite flexible in the sense that they impose no restrictions on the functional form of the ability–earnings relationships, the evolution of these relationships over time, or the way in which these relationships vary across our two levels of schooling. Indeed, we will argue that the application of these flexible methods enables us to find relationships in the NLSY data that may not have been uncovered using simpler linear regression models.

Using these methods, I find that the often-used assumption of a linear ability effect within time and education cells is rejected by the NLSY data. Further, this ability–earnings relationship varies across levels of schooling in a manner that can not be captured by including a simple ability–education interaction. Specifically, I find weak evidence of increasing returns to ability for those with at least some college education and strong evidence of diminishing returns to ability for those with 12 or fewer years of schooling. I argue that this result rationalizes why researchers have reported that the college wage premium is concentrated among the most able, as those with a college education continue to gain financially from added ability, while those without a college education are not compensated for moving farther into the right-tail of the ability distribution.

In addition to this, I follow Cawley *et al.* (2000) and Heckman and Vytlacil (2001) and investigate the extent to which we can ever really determine if returns to schooling are concentrated among the *most* able using data from the NLSY. In the absence of functional form assumptions, estimation of the returns to schooling for the most able requires data on the college and no-college wage outcomes of high-ability individuals. Loosely speaking, the most able individuals in the data select into college, while the least able do not, and thus estimates of the no-college wage for the most able must be critically dependent on arbitrary and potentially inappropriate parametric assumptions. Using a flexible regression model which does not impose such assumptions, I find that estimates of the college wage-premium for the most or least able are extremely inaccurate in the presence of ability-sorting into higher education. Although we may not be able to accurately determine the return to schooling for the *most* able, we do find that the wage premium is increasing for the *more* able over an ability support which is 'common' to both the college and no-college groups, and that simpler linear models understate this increase towards the upper limit of this common interval. I find that the growth in returns to schooling over the period 1984–94 occurred for

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

individuals at all points in the ability distribution. Further, I offer some weak evidence that this growth has followed a smoother, linear time-path for high-ability individuals than those of lower ability.

The outline of this paper is as follows. Section 2 presents the primary econometric model employed in this paper and discusses at a non-technical level a simple estimator for this model. Section 3 describes the data and presents the empirical results. The paper concludes with a summary in section 4.

II. The model

To focus attention on the difference between wages of the college educated and those without a college education, I partition the data into two levels of schooling which I denote as the 'college' and 'no-college' states. Individuals are assigned to the no-college state if they have completed 12 or fewer years of schooling and to the college state if they have at least 13 years of education.[9] To permit very general relationships between measured cognitive ability and our log wage outcomes within and across these two schooling states, I choose to estimate the following flexible *partially linear* log wage equations:

$$y_1 = \mathbf{X}\beta_1 + m_1(A) + \epsilon_1 \tag{1}$$

$$y_0 = \mathbf{X}\beta_0 + m_0(A) + \epsilon_0 \tag{2}$$

In the above y denotes the logarithm of hourly wages, $\mathbf{X}$ is a vector of characteristics affecting earnings in each state (such as experience and controls for local labour market conditions)[10] and A denotes our measure of cognitive ability. This ability measure will be a continuous test score standardized to have mean zero and unit variance which is constructed from the battery of ASVAB test scores in the NLSY. Finally, the '1' subscript will be used throughout the paper to denote parameters and variables associated with the college state and the '0' subscript to denote quantities in the no-college state.

[9]In Heckman and Vytlacil (2000), individuals are sorted into three groups: high school dropouts (education $<$12 years), high school graduates (education $=$ 12 years) and college graduates (education $=$ 16 years). I define the groups in this way to gather enough observations in each college state in each year, and also to enable nonparametric comparisons to be made between the two groups. Defining the college state to be those individuals with at least 16 years of schooling further restricts the ability support over which these comparisons can be made, as discussed below.

[10]Our empirical results were not significantly changed after further controlling for education within each of the two schooling groups. Specifically, the semiparametric and parametric estimates were not significantly altered after including a Bachelor's dummy in the college state (Ed $\geq$ 16) and a high school dummy in the no-college state (Ed $=$ 12). These 'sheepskin effect' variables were added to see if further nonlinearities in the schooling–earnings relationships affected key estimation results, but the impact of these added variables was found to be minimal.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

What separates (1) and (2) from most empirical specifications in this literature is the flexible manner in which we treat our ability variable, A. In particular, we do not impose any particular structure on the relationship between ability and log wages in each state, and simply let m_0 and m_1 be unknown *functions* mapping ability to log wages in the no-college and college states, respectively. The standard linear regression model would impose the prior restrictions $m_0(A) = \alpha_0^0 + \alpha_0^1 A$, $m_1(A) = \alpha_1^0 + \alpha_1^1 A$, and thus *bring to the data the assumption of linearity in ability.* In this paper, we do not bring anything to the data regarding the structure of the ability–log wage relationship, but instead, *let the data reveal to us the shapes of the ability–log wage relationships in each state.* Given our estimates obtained from this very flexible specification, we can then test if the ability log–wage relationships are, in fact, linear in each schooling state, or if they are equal across states. The flexibility afforded by this specification will also permit us to trace the college wage premium over the ability support, and thus enable us to examine the nature of the relationship between ability and the college wage gap without imposing particular functional forms.

The primary objects of interest in this analysis are the functions m_1 and m_0 and their derivatives. Note that the derivative functions $m_0{}'(A)$ and $m_1{}'(A)$ are marginal effects that are directly comparable with the regression coefficients that would be obtained from a simpler linear regression model. For example, if $m_0{}'$ and $m_1{}'$ are found to be constant over the support of A, then we would be back in the framework of the standard linear regression model, where the assumption of linearity imposes that the marginal effects are constant.

The estimates of the *functions* m_1 and m_0 are also useful in characterizing other quantities of interest, such as the expected log wage premium for given values of $\mathbf{X}$ and A:

$$\text{ATE}(\mathbf{X}, A) \equiv E(y_1 - y_0 \mid \mathbf{X}, A) = \mathbf{X}(\beta_1 - \beta_0) + m_1(A) - m_0(A)$$

We denote this expression as 'ATE' because of its similarity to the Average Treatment Effect parameter discussed in the programme evaluation literature.[11] For our purposes, this is simply a particular name for a parameter of great interest to us – the expected (or *average*) college log wage premium for an individual with known characteristics $\mathbf{X}$ and ability A.[12]

We also recognize that our college wage premium parameter ATE($\mathbf{X}$, A) is a function of *both* $\mathbf{X}$ and A, although what we are primarily interested in is

[11]For further discussion of treatment effects generally, see Heckman *et al.* (1998), Heckman and Vytlacil (2000) and Heckman, Tobias and Vytlacil (2001, 2002) among numerous others.

[12]We do not take up the issue of selection bias, but note that after controlling for ability directly, little correlation was found to remain between the unobservables in the college entry and log wage equations. In the absence of selection, we can consistently estimate the β parameters and functions m_1 and m_0 using the techniques discussed in this section.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

simply plotting the premium over the support of A. To this end, we require some method that will enable us to rid the dependence of the ATE parameter on the explanatory variables $\mathbf{X}$.

Ideally, we could accomplish this by obtaining estimates of $E(y_1 - y_0 \mid A) \equiv \text{ATE}(A)$ so that we can trace the expected log wage premium over the support of ability only without having to worry about values of the other explanatory variables. Thus, to fix ideas and purge ATE($\mathbf{X}$, A) of its dependence on the explanatory variables $\mathbf{X}$, we employ a method for averaging or 'integrating out' the effect of the $\mathbf{X}$ variables. The intuition behind the estimator we employ is to first imagine fixing ability equal to some particular value, say $A = A_0$. Then, consider taking an average of only the estimated ATE($\mathbf{X}$, A) values in the sample whose A values are 'close' to A_0. Proceeding in this way will enable us to select only those individuals whose Xs are associated with As close to A_0, so that we are essentially averaging the ATE($\mathbf{X}$, A) values over the distribution of $\mathbf{X} \mid A = A_0$. This will (under the requisite regularity conditions) produce a consistent estimate of $\text{ATE}(A_0) \equiv E(y_1 - y_0 \mid A = A_0) = E_{\mathbf{X} \mid A = A_0} E(y_1 - y_0 \mid \mathbf{X}, \quad A = A_0)$. This parameter *will not* depend in any way on the $\mathbf{X}$ values.

Although the technical details of this estimator are of secondary importance (and are discussed in greater detail in the appendix), the intuition behind our desire to use such an estimator is perhaps of primary importance. To come up with our 'best' prediction of the log wage premium at a given value of A, we would want to capture any 'indirect' information that the given value of A says about the values of $\mathbf{X}$. For example, the highest ability individuals in the data may forego participation in the labour market to acquire more schooling and thus we would see a negative sample correlation between ability and experience. As a result, we will be able to obtain a better marginal prediction of the log wage at high-ability values by taking into account the fact that high-ability individuals also tend to have relatively low experience. The estimator employed here automatically corrects for such correlation patterns in the data. Other approaches which might fix the values of $\mathbf{X}$ at sample means, for example, would be estimating a *conditional effect* given a particular value of $\mathbf{X}$, but would not fully purge the parameter of interest of its dependence on $\mathbf{X}$.[13]

A note on estimation

The power and contribution of the approach outlined in this paper is the flexibility of the employed models – we can explore the relationships between ability and log wages within and across our two schooling states without

[13]As a practical matter, however, we found that the resulting estimates using this method are very similar to those obtained after fixing $\mathbf{X}$ values at sample means, suggesting that the correlation is not practically significant.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

imposing any restrictive and potentially inappropriate functional form assumptions. This generality afforded by the model, however, also requires us to apply more advanced estimation techniques, as we are seeking to recover two unknown *functions* from the data (m_0 and m_1) rather than simply a finite number of regression parameters.

Fortunately, procedures for estimating models like the ones in (1) and (2) have received considerable attention in the recent literature, and it is worthwhile to explain the methods involved at a somewhat non-technical level. Typically these estimation procedures involve two stages. In the first stage, some technique is employed to 'remove' the unknown functions m_0 and m_1 from the model, enabling the researcher to then proceed to estimate the β parameters. In the second stage, estimates of β obtained from the first stage are used to obtain a *nonparametric* estimate of the unknown functions m_0 and m_1. We discuss below each of the two steps of this estimation process in greater detail.

First stage: estimation of β_i, $i = 0,1$

One estimator for the regression parameters β in this model exploits the assumed continuity of the m functions and uses the idea of differencing. To see how differencing might help us to estimate these parameters, consider two observations in our sample, say i and j, with virtually identical values of ability in the sense that $A_i \approx A_j$. Now, consider differencing equation (1) across these two observations to obtain:

$$\begin{aligned} y_{1i} - y_{1j} &= (\mathbf{X}_i - \mathbf{X}_j)\beta_1 + m_1(A_i) - m_1(A_j) + \epsilon_{1i} - \epsilon_{1j} \\ &\approx (\mathbf{X}_i - \mathbf{X}_j)\beta_1 + \epsilon_{1i} - \epsilon_{1j} \end{aligned} \tag{3}$$

The first line above simply differences our regression model across these two observations, while the second line exploits the assumed continuity of m_1 and notes that for $A_i \approx A_j$, we can assume $m_1(A_i) \approx m_1(A_j)$. Thus, by taking differences of the data across 'adjacent' ability values, we can approximately difference out the unknown function m_1 so that only the β_1 parameters remain to be estimated.

A simple extension of these arguments leads us to the following two-step differencing procedure for obtaining estimates of the β_i, $i = 0,1$:

1. Within each of the two schooling states, sort the data by ascending values of A.
2. Take first-differences of the sorted data and estimate β_i by an OLS regression of the differenced $y_i's$ on the differenced $\mathbf{X}$s in each state.

The consistency and asymptotic normality of this simple differencing estimator are described in Yatchew (1997). The simple differencing technique

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

outlined above is not new to this literature and has been used in past work by Ahn and Powell (1993) among others. In this paper, we use a slightly more general estimator and employ the use of higher-order differencing with optimal differencing weights to improve the efficiency of this simple first-differencing estimator. The intuition behind this generalized differencing estimator is exactly the same as simple first-difference estimator, although Yatchew (1997) shows that one can increase efficiency by increasing the order of differencing. For this reason, we choose to use optimal tenth-order differencing,[14] although we note that both longer and shorter orders of differencing produced highly similar estimates of the β parameters.

Second stage: estimation of m_1 and m_0

To this point, we have only discussed estimation of the linear part of our partially linear model and have seen how differencing provides a strategy for consistent estimation of β_1 and β_0. In this section, we discuss in general terms how the functions m_0 and m_1 can also be estimated given our first stage estimates of β_0 and β_1.

We begin by focusing on estimation of the college state equation in (1) and note that if the regression parameter β_1 were known, we could simply move the linear part of (1) to the left-hand side and form a new dependent variable:

$$y_{1i} - \mathbf{X}_i\beta_1 = m_1(A_i) + \epsilon_{1i} \tag{4}$$

or

$$\tilde{y}_{1i} = m_1(A_i) + \epsilon_{1i}, \tag{5}$$

where $\tilde{y}_{1i} = y_{1i} - \mathbf{X}_i\beta_1$. The above equation is in the framework of a standard univariate *nonparametric* regression problem. The model in (5) and the approach for estimating it are called *nonparametric* because we are not parameterizing the function m, but instead, treat it as a completely unknown quantity without imposing additional restrictions.

To estimate the functions m_1 and m_0, we will obtain point estimates of them at a series of different grid points spaced throughout the ability support and then piece these point estimates together. The formal technique we use to obtain point estimates of the m_i at the various grid points is called *local linear regression*. As the name suggests, the idea behind local linear regression is to fit a linear regression locally around a neighbourhood of a particular grid point A_0 to provide a point estimate of $m_i(A_0)$, $i = 0,\ 1$. This local regression is carried out by first weighting all the data points in the sample according to

[14]See Yatchew (1997, 1998, 1999) for more on this technique and for explicit values of the differencing weights at various orders of differencing.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

their 'distance' from A_0 – observations with A_i values 'close' to A_0 will receive high weight in our objective function while points far away from A_0 will receive little or no weight. By fitting a series of local regressions, we are able to recover the shape of the unknown functions even if they fluctuate rather erratically. Investigating the behaviour of m_i locally *lets the data reveal to us the shape of the regression functions*, while a standard parametric approach which might impose $m_i(A) = \alpha_i^0 + \alpha_i^1 A$ necessarily *brings to the data the shape of the regression functions*.

The mechanics behind the weighting scheme involved in local linear regression are carried out through the choice of a *kernel function* and a *bandwidth* or *smoothing parameter*. These quantities generally control the size of the local neighbourhood around the grid point A_0 as well as the overall 'smoothness' of the resulting estimates of m. Most important in this process is the choice of bandwidth, and given this sensitivity to bandwidth choice, we use the rule-of-thumb bandwidth selector of Fan and Gijbels (1996; pp. 110–112) to guide our choice of smoothing parameter.

Of course, the nonparametric regression method described in this section proceeded under the assumption that the β parameters are known. A natural way around this problem is to use the β estimates obtained from the first stage of this estimation process to construct a new dependent variable [i.e. instead of $\tilde{y}_1 = y_1 - \mathbf{X}\beta_1$ in (5), use $y_1 - \mathbf{X}\hat{\beta}_1$]. This choice turns out to perform quite well, and in fact, asymptotic SE associated with our estimates of m_0 and m_1 will not be affected by the first-stage estimation of the linear regression parameters β_0 and β_1.

Here we have only offered a brief and intuitive explanation of the procedures behind nonparametric regression. A more detailed description at the introductory level that contains both applications and generated data experiments can be found in DiNardo and Tobias (2001). More comprehensive and advanced treatments of kernel regression methods generally can be found in Härdle (1990), Fan and Gijbels (1996), and Blundell and Duncan (1998), among others.

III. The data and empirical results

The data used in this analysis are taken from NLSY. The NLSY is a rich panel data set containing information on the earnings, education, occupation and other characteristics of young men and women aged 14–22 years in 1979. The version of the NLSY used in this analysis contains annual survey data from 1979–94.

The sample is restricted to white males not currently enrolled in school reporting earnings between \$1 and \$100 per hour in the given year. The regressors (variables in $\mathbf{X}$) include an intercept, indicators for residence in the

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

northeast and south, potential labour market experience and its square, a measure of the local unemployment rate and an indicator for residence in an urban area. We control for education by partitioning the data into two schooling groups and analysing each group separately. The remaining explanatory variable and the primary source of interest in this analysis is the ability measure, which we denote as g. This ability measure is the first principal component of the 10 standardized tests in the ASVAB residualized on age, which is then standardized to have mean zero and unit variance.[15] The 10 component tests comprising the ASVAB battery are general science, arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and shop information, mathematics knowledge, mechanical comprehension and electronics information. In this paper, I use g as my ability measure, which is the scalar variable explaining the largest amount of variation (nearly 55 percent) in the 10 component tests. Of course, higher-order principal components can also be used to explain variation in earnings, as discussed in Cawley *et al.* (1997). In this paper, I use g as my overall measure of ability, and defer the analysis of alternate ability measures and the use of higher-order principal components as the subject for future work.[16]

Individuals are assigned to the college state if they have completed at least 13 years of schooling by the given year and are otherwise assigned to the no-college state. The earnings variable used is reported hourly wages in the given year, which are measured in 1990 dollars. After excluding observations for the reasons listed above, we obtain a total of 12,857 person-year observations, 5312 of which fall into the college state and 7545 into the no-college state.

Empirical results

In this section, we look at the data to address the following questions: (1) What do the functions mapping ability to earnings look like? Are they linear within education and time cells, or constant across our two schooling groups?

[15]As individuals in the sample vary in age at the time of the test, each of the 10 tests is first residualized on age and g is defined as the first principal component of these standardized residuals. Similar results are obtained when the test scores are residualized on age and a college indicator. Armed Forces Qualifying Test (AFQT), an average of four of the 10 component tests, has often been used as a measure of cognitive ability in past work using NLSY data. The ability measure used here, g, is highly correlated but more general than AFQT, as it incorporates information in all 10 of the component tests.

[16]The use of g and related test scores constructed from the ASVAB battery has a considerable precedent in empirical work. See, for example, Blackburn and Neumark (1993), Herrnstein and Murray (1994), Cawley *et al.* (1997), Heckman *et al.* (2001) and Heckman and Vytlacil (2001). Blackburn and Neumark (1993) also investigate the problem of measurement error when using AFQT score and conclude that measurement error does not seem to be a significant problem with this ability measure, but demonstrate that measurement error was problematic in past work when other ability measures such as IQ were used.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(2) Are standard parametric models adequate for describing the ability–earnings relationships or assessing the effect of ability on returns to schooling? (3) In the presence of ability-sorting into higher education, can we accurately determine the return to schooling for the *most* able? (4) What can our flexible analysis tell us about the relationship between ability and the return to schooling over an 'interior' ability support? Finally, (5) How have the returns to schooling evolved over time for individuals of differing ability?

The Shape of the ability–earnings relationships

To begin, to address these questions, I first explore the relationship between ability and log wages across two levels of schooling, which I have denoted as the 'college' and 'no-college' states. Figures 1 and 2 present results of this exploratory data analysis, and plot estimates of the *derivatives* $m_0'(A)$ and $m_1'(A)$ and their associated SE after pooling observations over time periods. Recall that our ability measure A is standardized to have mean zero and unit variance so that $A = 1$, for example, corresponds to an individual 1SD above the mean of the ability distribution.

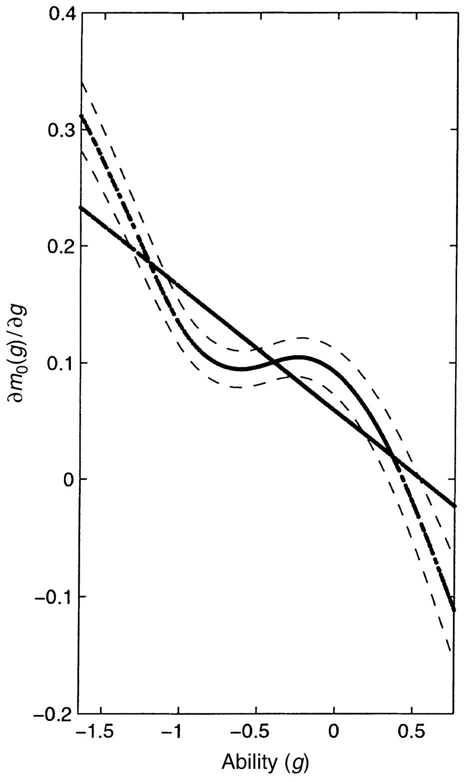

Figure 1. No-college state (Ed ≤ 12)

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

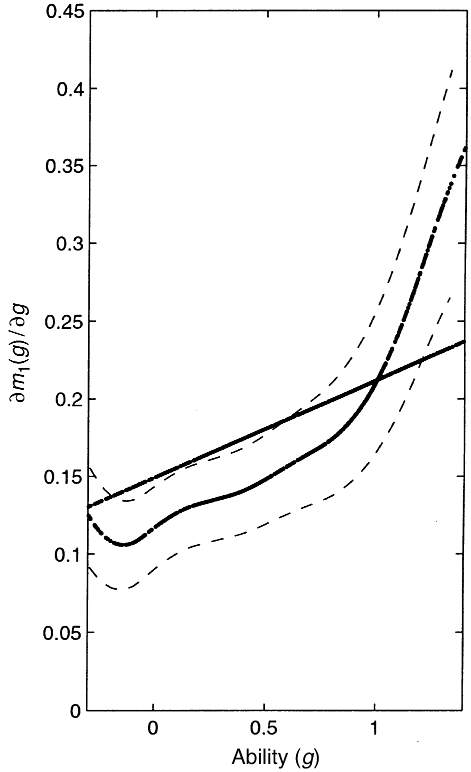

Figure 2. College state (Ed > 12)

I obtain plots of the marginal effects rather than the conditional mean functions m_0 and m_1 as the marginal effects are directly comparable with the coefficient on ability obtained from a regression of log wages on ability.[17] The numbers on the vertical axis (multiplied by 100) give the percentage change in wages resulting from a 1SD jump in the ability distribution, and the flexibility of the partially linear model allows these percent changes to vary at all points throughout the ability support. As discussed in section 2.1, estimates of the β parameters are obtained using Yatchew's (1997) optimal tenth-order differencing estimator for the partially linear regression model. In the second stage (see section 2 for a description), estimates of the derivatives of m_0 and m_1 are obtained using local quadratic regression.[18] Results are obtained after pooling

[17]Further, estimates of m_1 and m_0 necessarily include the intercept term.

[18]A biweight kernel is used and the bandwidth is chosen optimally using a slightly oversmoothed version of the pilot bandwidth selector of Fan and Gijbels (1996). The general trends discussed here are not sensitive to the choice of bandwidth, although the smoothness of the nonparametric point estimates is sensitive to the bandwidth choice. Local quadratic regression is used because of the well-known preference for odd-order fits, as the objects of interest here are the first derivatives. Pointwise SE are obtained using the bootstrap.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

observations across time periods and interacting the variables in **X** with year dummies.

Figures 1 and 2 are quite revealing. First, and perhaps most importantly, the graphs provide evidence of diminishing marginal returns to ability for those in the no-college state (Figure 1) and increasing returns to ability for those in the college state (Figure 2). The dashed lines in the figures are large-sample 95 percent pointwise confidence intervals obtained using the bootstrap. These SE are reasonably small and suggest rather convincing evidence of nonlinearities that vary across the schooling groups. To further emphasize the importance of the nonlinearities, suppose that the models in (1) and (2) were actually linear in ability. In this case, the marginal effects would be constant and thus the partially linear estimates in Figures 1 and 2 should resemble a horizontal line. Clearly, the results in the figures are not flat and we can not draw a horizontal line within the SE bars over the ability support. These results suggest strong evidence of a nonlinear relationship between ability and log wages and this relationship clearly varies across college states.

Taken together, the plots in Figures 1 and 2 help to explain why researchers have reported that returns to education have been concentrated among the most able. From the graphs, we see that the most able individuals are continually (and increasingly) compensated for their ability in the college state, while they are not compensated (after a certain point) in the no-college state.[19] This suggests that individuals without a college education do not benefit from moving into the far right tail of their ability distribution, and conversely, individuals with a college education benefit significantly from moving into this far right tail of the ability distribution.

For the sake of comparison, Figures 1 and 2 also provide estimates obtained from a restricted model that is capable of reproducing the main points suggested by our semiparametric point estimates. Equations (1) and (2) were also estimated using a restricted quadratic specification of ability (i.e. $m_i(A) = \alpha_i^0 + \alpha_i^1 A + \alpha_i^2 A^2,\ i = 0, 1$) and the (linear) marginal effects associated with this specification are also plotted in Figures 1 and 2. If the previous interpretation of the partially linear derivative estimates was correct, we would expect to see a positive and significant coefficient associated with the quadratic term of ability in the college state (increasing returns) and a negative and significant term in the no-college state (diminishing returns). This is indeed the case. The coefficients (and heteroscedasticity-consistent t-statistics) for the linear and quadratic terms associated with the ability variables in this regression are 0.149 (13.45) and 0.031 (3.40), respectively, in the college state and 0.058 (6.24) and −0.053 (−10.59), respectively, in the no-college state.

[19]In fact, regressions with a quadratic component in g suggest that the marginal returns to ability for those with 12 or fewer years of education and g values near 3/4 of an SD above the sample mean are nearly zero in all years studied.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

These results again reveal compelling evidence of nonlinearities in the ability–log wage relationship that vary across levels of schooling.

While the results presented in Figures 1 and 2 have used flexible econometric techniques to reveal an interesting empirical finding, they are also restrictive and may not be fully satisfactory. These results are obtained by pooling observations across time periods, which precludes any possible changes in the ability–earnings relationship and the college wage premium over time. We can imagine several reasons while the pooling assumption is not warranted: (1) Ability is known to be correlated with earnings over the life cycle (e.g. Griliches and Mason, 1972). High-ability individuals are likely to select higher education and thus experience similar or even lower earnings than individuals with less schooling when they are young. This suggests that the ability–earnings relationship may differ by age across college states, and given the structure of the NLSY data, these relationships will also differ by survey year.[20] (2) Previous empirical work has documented changes in the ability–earnings relationship and the college wage premium over time, particularly during the 1980s. Herrnstein and Murray (1994) and Murnane *et al.* (1995) show that the return to education has risen during a similar period of study and argue that a substantial fraction of the return to education can be explained by a rising premium paid to ability. Blackburn and Neumark (1993) report the significance of an ability–education–time interaction[21] and conclude that the increase in returns to schooling over time has been experienced primarily by the most able. Taber (2001) provides some evidence that a rising demand for unobserved ability may be partially responsible for the growing US college wage premium, and Heckman and Vytlacil (2001) find that returns to schooling have risen over time primarily for those in the highest ability quartile. Heckman and Vytlacil (2001) also test and reject the widespread use of linear time and age effects, and reject the use of linear time effects within education and ability cells.

The results of these studies suggest that (a) the pooled model is too restrictive in the sense that it does not enable us to trace the college wage premium across years and investigate the possibly changing role of ability in explaining this wage premium, and (b) generalizing this model by simply adding time–education, time–ability, or time–education–ability interactions may not be empirically justified, as documented in Heckman and Vytlacil (2001) who also use NLSY data. To this end, I free the restriction that the functions m_1 and m_0 are time-invariant and obtain results separately for each year from 1984–94. By analysing results separately for each year, I am not

[20]In the NLSY, a group of individuals aged 14–22 years is followed from 1979. Thus, everyone in the sample ages 1 year at each consecutive interview, so that time and age are inevitably intertwined. See Heckman and Vytlacil (2001) for further discussion of this issue.

[21]Specifically, they report the significance of this interaction using an ‘academic’ ability measure.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relying on any functional form assumptions to describe the evolution of the ability–earnings relationship or college wage premium over time. In this way, I can obtain year-by-year estimates of the college wage premium and plot this return over the ability support without assuming potentially inappropriate functional forms.

To begin this year-by-year investigation, we present in Table 1 the estimated coefficients using a quadratic specification of ability, running separate regressions for each of the years 1984–94. Given our results in Figures 1 and 2, we would expect to see negative coefficients on the no-college state quadratic terms and positive coefficients on the college-state quadratic terms. The results in Table 1 are highly consistent with our pooled estimation results and show that the coefficient on the quadratic ability term in the no-college state is always negative and is statistically significant at the 1 percent level in 10 of 11 years. In the college state, the coefficient on the quadratic term is positive in 9 of 11 years and statistically significant at the 10 percent level in 5 of 11 years and at the 5 percent level in 4 of 11 years. Again, this provides year-by-year evidence of diminishing returns to ability for those with 12 or

TABLE 1

Coefficients and heteroscedasticity-consistent t-statistics using a quadratic specification of g

		College state		*No-college state*	
Year	*Variable*	*Coeff.*	*t-Stat*	*Coeff.*	*t-Stat*
1984	Linear (g)	0.081	1.28	0.065	2.51
	Quadratic (g^2)	0.082	1.76	−0.054	−3.12
1985	Linear (g)	0.182	4.38	0.004	0.118
	Quadratic (g^2)	−0.006	−0.147	−0.052	−3.64
1986	Linear (g)	0.169	2.93	0.103	3.69
	Quadratic (g^2)	0.074	1.66	−0.047	−2.70
1987	Linear (g)	0.116	2.52	0.065	2.13
	Quadratic (g^2)	0.016	0.418	−0.049	−2.99
1988	Linear (g)	0.162	4.85	0.051	1.67
	Quadratic (g^2)	0.050	1.61	−0.051	−2.84
1989	Linear (g)	0.159	4.30	0.054	1.75
	Quadratic (g^2)	0.053	1.92	−0.060	−3.53
1990	Linear (g)	0.149	4.14	0.070	2.26
	Quadratic (g^2)	0.031	1.02	−0.043	−2.59
1991	Linear (g)	0.169	5.31	0.068	2.23
	Quadratic (g^2)	0.018	0.714	−0.068	−4.00
1992	Linear (g)	0.103	3.27	0.053	1.84
	Quadratic (g^2)	0.037	1.40	−0.060	−3.55
1993	Linear (g)	0.112	3.70	0.074	2.54
	Quadratic (g^2)	0.054	2.14	−0.031	−2.10
1994	Linear (g)	0.186	4.51	0.043	1.31
	Quadratic (g^2)	−0.026	−0.708	−0.054	−2.98

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

fewer years of schooling and weaker evidence of increasing returns to ability for those with at least some form of college education.

Tests against parametric alternatives

The results of our exploratory data analysis suggest that models imposing linearity in ability or equality of the ability–earnings relationships across levels of schooling are not supported by the NLSY data. In this section, we formally test these hypotheses in our partially linear framework using the tests described in Yatchew (1997, 1999). Blackburn and Neumark (1993) obtained estimates of the college wage premium by writing earnings as a parametric function of time, ability, education, and interactions of these three variables. Their specification implies linear ability–earnings relationships within education and time cells and also imposes that the marginal effect of ability on wages is linear in education within a given time cell. These restrictions are not capable of producing the relationships found in the beginning of this section and thus we seek to test in a formal way if these often-invoked parametric assumptions are appropriate.

I begin by testing the partially linear model against two parametric alternatives: a model without an ability term, $(m_i = \alpha_0^i)$ and another with a linear ability component $(m_i = \alpha_0^i + \alpha_1^i A)$ for each year over the period 1984–94.[22] The test statistics associated with these null hypotheses are provided in Table 2. With the exception of the college state in 1992, I strongly reject the

TABLE 2

Test statistics for alternate restrictions

	College state		*No-college state*		
Year	$H_0 : m_1 = \alpha_0^1$	$H_0 : m_1 = \alpha_0^1 + \alpha_1^1 A$	$H_0 : m_0 = \alpha_0^0$	$H_0 : m_0 = \alpha_0^0 + \alpha_1^0 A$	$H_0 : m_0 = m_1$
1984	3.82	2.27	8.30	4.15	11.62
1985	5.42	0.902	6.12	3.83	6.11
1986	7.77	3.11	9.35	2.60	10.71
1987	2.81	1.07	6.93	2.81	10.10
1988	6.65	1.97	5.67	1.67	11.32
1989	4.95	0.134	7.77	2.92	13.24
1990	3.39	−0.026	5.70	1.08	8.05
1991	3.21	−0.68	10.08	3.42	11.47
1992	1.34	−0.576	8.93	3.34	8.40
1993	3.12	0.981	5.85	1.79	9.83
1994	4.09	0.399	5.80	1.92	9.22

[22] Yatchew (1997) shows under the null hypothesis that the function m takes some given parametric form, $(mn)^{1/2}(s^2_{\text{diff}} - s^2_{\text{rest}})/s^2_{\text{diff}} \sim N(0,1)$ where s^2_{diff} is the estimated variance parameter obtained by optimally differencing out the unknown regression function, s^2_{rest} is the estimated variance parameter when using the restricted parametric model and m denotes the order of differencing. Note that nonparametric estimation is not required in order to implement these tests.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

null hypothesis of no ability effect ($H_0 : m_i = \alpha_0^i$) for all years and for both education groups. Thus, our ability measure is an overwhelmingly important factor in explaining variation in log wages and clearly needs to play some role in our regression models.

As for the linearity tests, for those with 12 or fewer years of schooling, I reject (at the 10 percent level) the null hypothesis of a linearity (i.e. $H_0 : m_0 = \alpha_0^i + \alpha_1^i A$) for all years except 1990. For those in the college state, I can only reject linearity (at the 5 percent level) in 1984, 1986 and 1988. These results are again consistent with those obtained from the quadratic estimates, which show clear evidence of nonlinearities for those with 12 or fewer years of schooling and weaker evidence of nonlinearities for those with some amount of college education.

In addition to the tests against parametric alternatives, we also test for the equality of regression functions ($H_0 : m_1 = m_0$) using the method described in Yatchew (1999).[23] If these functions are equal, then plots of $E(y_1 - y_0 \mid A)$ will be constant over A, implying that the college wage premium is independent of ability. As shown in Table 1, we strongly reject the null hypothesis of equal ability relationships across college states for all years studied and thus conclude that ability plays an important role in explaining the college wage premium.

To summarize our answers to questions (1) and (2) posed at the outset, we find very strong evidence of nonlinearity in the ability–earnings relationship for those with fewer than 12 years of schooling and weak evidence of nonlinearity for those with at least 13 years of education. These relationships are revealed from our semiparametric estimation in Figures 1 and 2, from our restricted estimation using a quadratic ability specification and also from our general specification tests against parametric alternatives. We also find very strong evidence that the ability–log wage relationships vary across our two schooling groups. These results suggest that estimates of the college wage premium must be based on models that are flexible enough to capture these features of the data. Models that do not allow heterogeneity in the ability–earnings relationships across college states, or assume linearity within education and time cells are likely to give misleading and biased estimates of the returns to schooling.

The college wage premium and the identification problem

In Figure 3 we plot $E(y_1 \mid A)$ (the expected log wage of those in the college state), $E(y_0 \mid A)$ (the expected log wage of those in the no-college state) and

[23]See Yatchew (1999), p. 22. This test requires one to obtain estimates of the variance parameters for the 'within' and pooled models, and also to determine the probability that adjacent observations in the sorted pooled data come from the same population (here, college or no-college states).

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

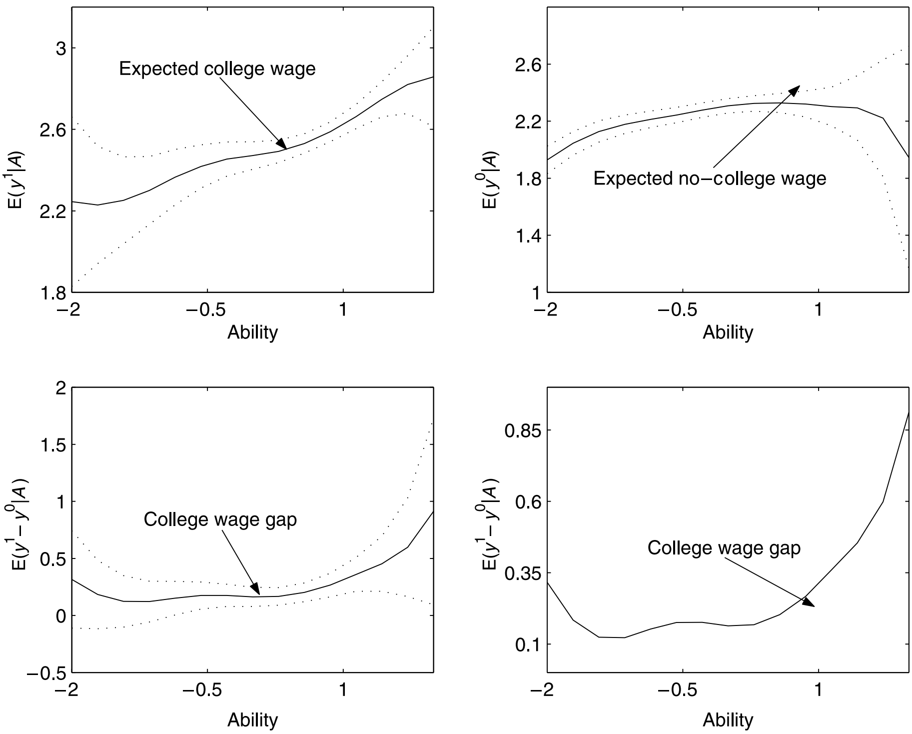

Figure 3. Estimated expected log wages in the college state [$E(y_1|A)$], no-college state [$E(y_0|A)$], and the expected log wage premium [$E(y_1 - y_0 \mid A)$] using 1990 NLSY data

$E(y_1 - y_0 \mid A)$ (the college log wage premium) over the interval [−2,2] using 1990 data. We obtain these plots by integrating out the effect of the linear **X** characteristics, as discussed in section 2 and further discussed in the appendix. The graphs also contain large-sample 95 percent confidence intervals. These confidence intervals are constructed using an optimally-smoothed higher-order local cubic regression to characterize the bias and variance of the local linear regression estimator.[24]

The plots reveal several interesting results. First, we see that expected log wages are increasing over the full ability support for those in the college state (upper left), while they turn down after a certain point for those in the no-college state (upper right). These results are consistent with those obtained in Table 1. The ability–earnings relationship in the college state appears roughly linear, while the no-college relationship resembles a quadratic function. Secondly, we see that the expected log wage estimates

[24]See Fan and Gijbels (1996), pp. 113–120.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

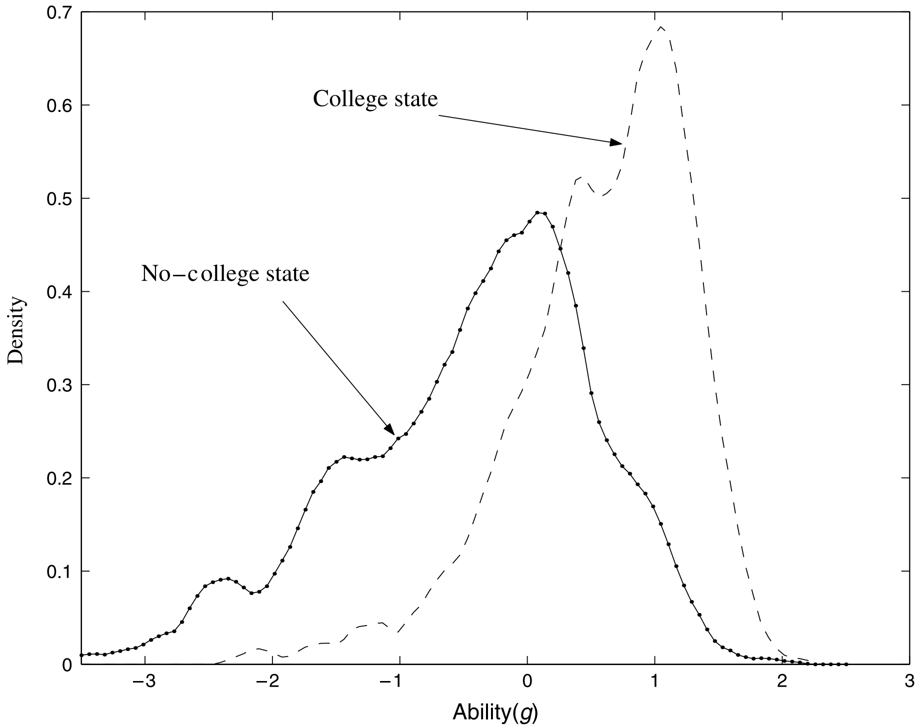

Figure 4. Density of g across college states

are very imprecise at the left tail of the ability distribution for those in the college state (upper left) and at the right tail of the ability distribution for those in the no-college state (upper right). This imprecision in the tails is a direct consequence of *ability-sorting into higher education* (e.g. Cawley *et al.*, 2000; Heckman and Vytlacil, 2001): individuals with at least some college education tend to have much higher ability values than those not selecting higher education, and thus the scarcity of high-ability no-college individuals and low-ability college individuals leads to imprecise estimates over these regions.

Figure 4 demonstrates this striking ability-sorting problem and plots the density of our ability measure, g, across the college and no-college states.[25] What we see in Figure 4 is that the distribution of ability is shifted to the right in the college state relative to the distribution in the no-college state. For ability values <0, we are essentially in the left tail of the college-state ability distribution and for ability values >1 we are in the right tail of the no-college

[25]These density estimates are obtained using a biweight kernel using Silverman's (1986) optimal bandwidth formula. Results are obtained using 1990 data and are extremely similar to results obtained using g values for different years.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

state ability distribution.[26] This feature of the data will limit our ability to precisely estimate wages for college graduates at the left tail of the ability distribution and wages for those in the no-college state at the right tail, as evidenced in the top row of Figure 3.

The modest overlap of the ability distributions across states certainly casts doubt on the conventional wisdom that that the recent rise in returns to education has been concentrated among those with highest ability. As argued above, the scarcity of very high-ability individuals in the no-college state makes it difficult for us to identify (in the absence of any arbitrary functional form assumptions) the returns to schooling for the most able. Indeed, this is evident in Figure 3. At the left and right tails of the ability distribution, our point estimates of the log wage gap are very imprecisely estimated (Figure 3, lower left). For g values >1SD above the mean, our confidence intervals quickly spread out, primarily because of the uncertainty in estimating the no-college expected wage at high ability levels. At this right tail, we cannot be sure of the magnitude of the wage premium, although our point estimate is clearly increasing in ability over this region. Over the ability support which is 'interior' to both states, the log wage premium is much flatter and the confidence intervals are much tighter and bounded away from zero. If our question is to identify the log wage premium for the most or least able, we cannot know with reasonable precision using data like the NLSY without extrapolating by making potentially inappropriate functional form assumptions.

Given these results, we proceed to estimate our expected log wage premium $E(Y_1 - y_0 \mid A)$, denoted ATE(A), over an ability support which is 'interior' to each group, and thus can be estimated reasonably accurately without resorting to functional form assumptions. To this end, we let A_c^p be the p^{th} percentile of the ability distribution in the college state and define A_{nc}^p similarly for the no-college state. We then choose our ability support to be the interval $[\underline{A}, \bar{A}]$ where

$$\underline{A} \equiv \max\{A_c^5, A_{nc}^5\}$$
$$\bar{A} \equiv \min\{A_c^{95}, A_{nc}^{95}\}$$

In all cases, the largest of the 5th percentiles is found in the college state and the smallest of the 95th percentiles is in the no-college state. Roughly, this rule leads us to estimate the log wage gap over the interval $[-1,1]$.

We estimate the expected log wage premium semiparametrically for all years (1984–94) over the ability supports defined above. In Figure 5, we

[26]Specifically, the 10th and 90th percentiles of the g distribution are -0.29 and 1.35 in the college state (respectively) and -1.76 and 0.74 in the no-college state.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

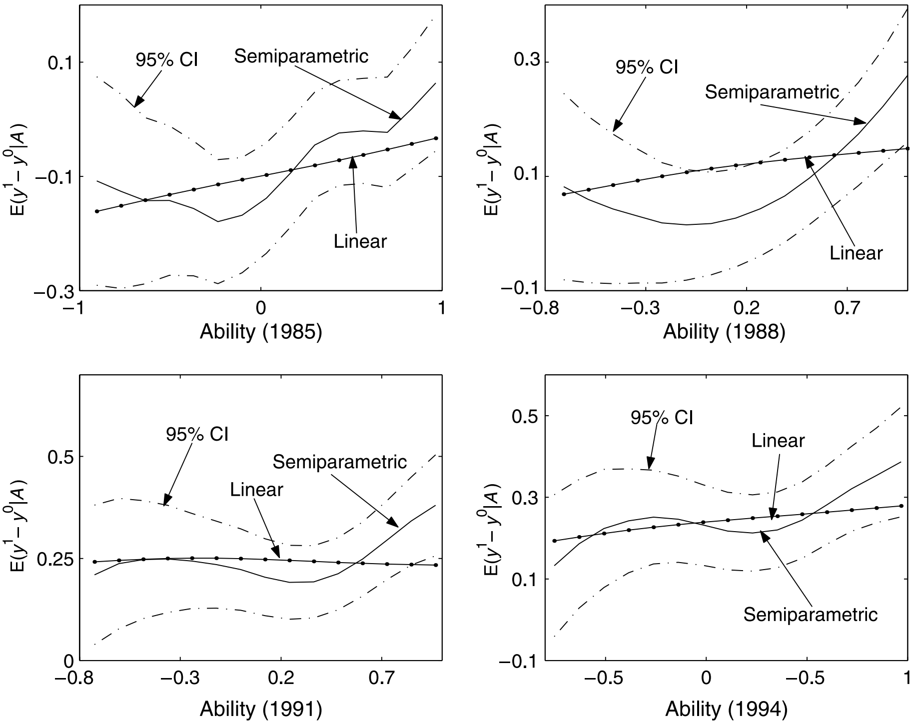

Figure 5. Linear and semiparametric estimates of ATE(A) ≡ $E(y^1 - y^0 \mid A)$: 1985, 1988, 1991 and 1994

present results of this analysis for data spaced apart by 3-year intervals: 1985, 1988, 1991 and 1994, and provide results for the remaining years in Figures 7 and 8. For the sake of comparison with more familiar and widely-used methods, we also carry along estimates of $E(y_1 - y_0 \mid A)$ when using a linear specification of ability in both college states: $m_i(A) = \alpha_0^i + \alpha_1^i A$.

From Figure 5, we first see that the wage gap is generally increasing over the ability support for all years studied. This suggests that returns to schooling are indeed concentrated among the *more* able, even after we have limited ourselves to an ability support which is ‘interior’ to both college states and freed ourselves of parametric assumptions. When comparing the semiparametric results to those obtained from the linear model, we find that the linear estimate consistently understates the wage gap towards the right tail of the ability support, although the linear point estimates fall within our large-sample 95 percent confidence intervals. This nonlinearity at the right tail is a consequence of diminishing returns to ability for those with 12 or fewer years of schooling. While the simple linear model is constrained to predict that $E(y_1 - y_0 \mid A)$ is linear in A, the diminishing returns to ability in the

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

no-college state implies that ATE should be increasing at an increasing rate as we move into the right tail of the ability distribution. This is indeed suggested by our semiparametric point estimates, but cannot be captured by our linear estimates. These results potentially suggest that returns to schooling are concentrated among the *more* able to an even greater extent than has been suggested by previous analyses that impose linear ability relationships within time and education cells. Using 1994 outcomes, for example, we find that an individual 1SD above the mean of the ability distribution gets a 40 percent increase in hourly wages from selecting the college state, while the linear model predicts an increase of slightly less than 30 percent.

Ability and the growing college wage premium

In question (5) proposed at the outset of this section, we sought to determine how returns to schooling have changed over time for individuals of differing ability. In Figure 6, we provide an answer to this question. We consider three hypothetical individuals with three different ability values ($A = -1$, 0, 1, denoting an individual of average ability, and individuals 1SD below and 1SD above the mean of the ability distribution) and trace the estimated log wage gap over time for these three individuals.

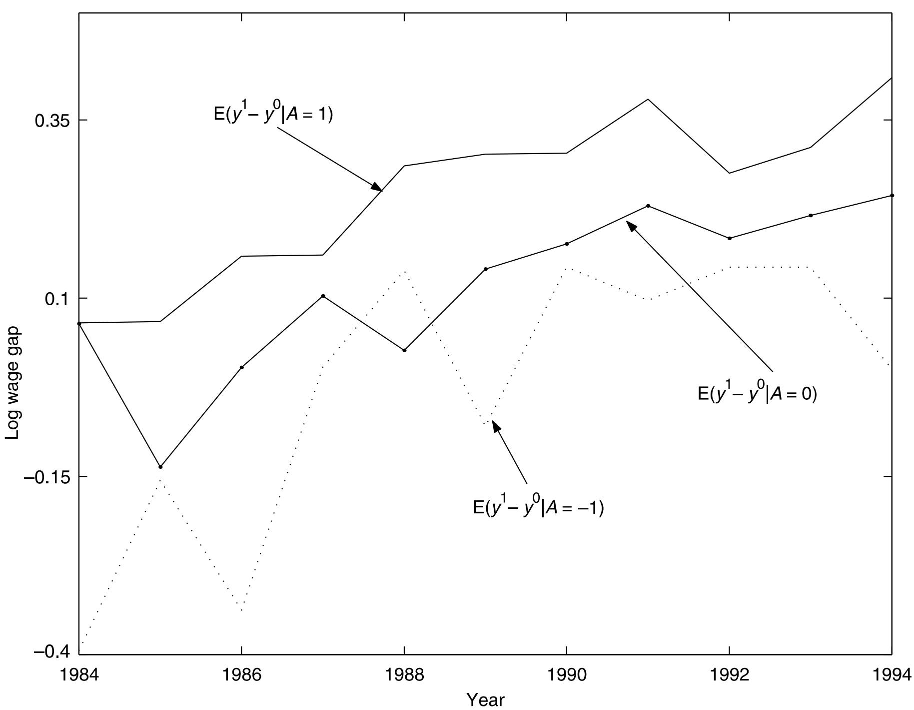

Figure 6. Semiparametric estimates of ATE(A) over time for $A = -1$, 0, 1

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

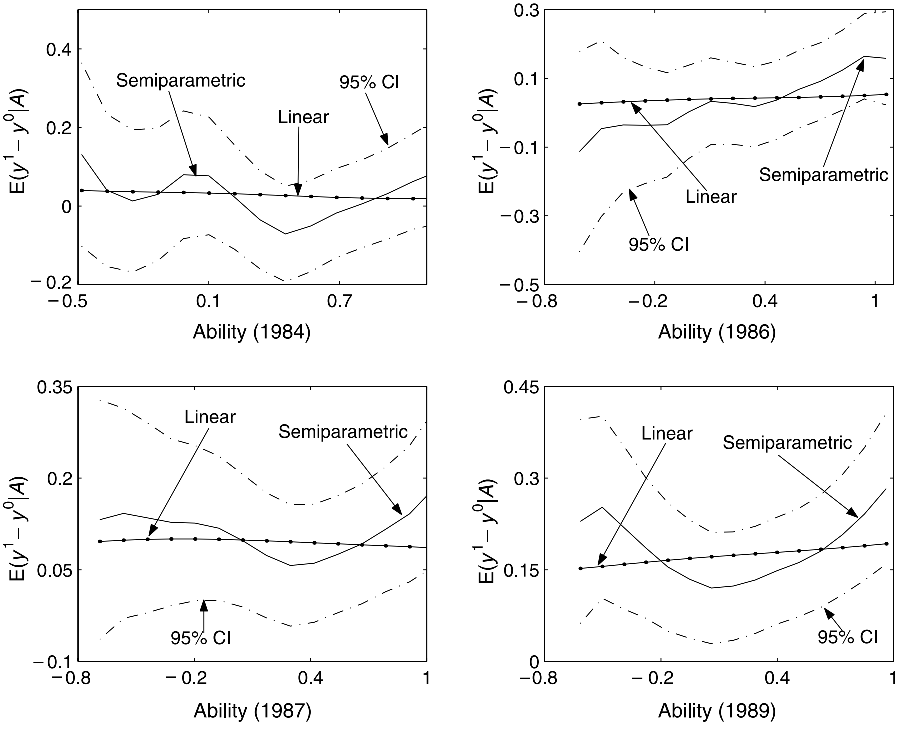

Figure 7. Linear and semiparametric estimates of ATE(A) ≡ $E(y^1 - y^0 \mid A)$, 1984, 1986, 1987 and 1989

The results of this analysis shown in Figure 6 reveal several interesting patterns. First, as a general rule, the college log wage premium is greater for higher-ability individuals than their lower-ability counterparts for each year studied, which is consistent with the positive slopes found in Figures 5, 7 and 8. Secondly, the trends are upward-sloping for each of the three ability groups, indicating a rising wage premium over time for individuals of all ability types. From Figure 6, we see that the increase in returns to schooling over the period 1984–94 seems to occur at all three ability levels ($A = -1$, 0, 1), and are largest in magnitude for the most able individuals. For the lowest ability group, the wage premium remains consistently positive only for the latter years of the sample: 1990–94, but is generally increasing (albeit erratically) throughout this period. This figure also provides some suggestive information that the time-path of the log wage premium may be less variable for the highest-ability individuals than the lowest-ability individuals. Estimates of $E(y^1 - y^0 \mid A = 1)$ appear to be 'smooth' and consistently rising, while the graph of the wage gap for the lowest ability group ($A = -1$) is quite jumpy from year to year.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

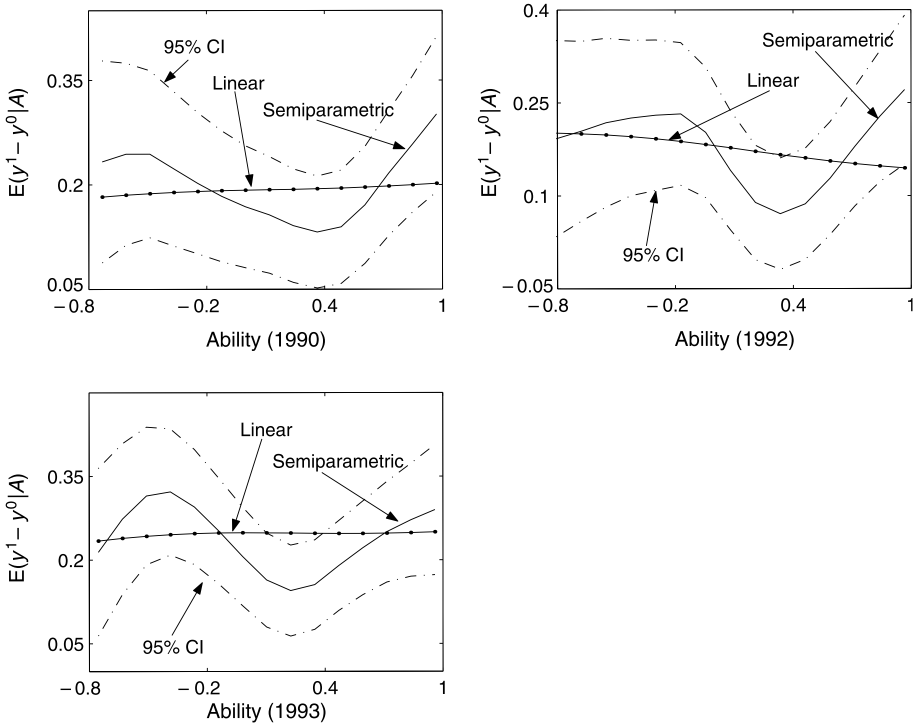

Figure 8. Linear and semiparametric estimates of ATE(A) ≡ $E(y^1 - y^0 \mid A)$, 1990, 1992 and 1993

To further explore the issue, we regressed the year-by-year point estimates for the $A = 1$ and $A = -1$ groups in Figure 6 on a constant and a linear time trend. The regression for the high-ability group produced an R^2 of 0.81 while the regression for the low-ability group ($A = 1$) yielded an R^2 of 0.55. This suggests that a model which imposes a linear growth path for the log wage premium over time fits rather well for the high-ability group ($A = 1$), but explains only half of the variation over time for the lowest-ability group. Thus, the most able not only experience the highest returns to schooling but may also face comparably less uncertainty in the linear growth of these returns over time than those of lower ability. This last point has not been explored in detail in previous work and provides an interesting direction for future research.

IV. Conclusion

Are returns to schooling concentrated among the most able? In this paper, I have argued that this might not be the appropriate question to ask, or at least, the data will not enable us to accurately answer it without making strong functional form assumptions. In data sets like the NLSY, the most able individuals get into college with very high probability, and thus we cannot accurately estimate the

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

no-college outcomes of high ability individuals without resorting to some type of potentially inappropriate parametric assumption.

In this paper I explored the ability–earnings relationships using flexible econometric techniques that do not require such functional form assumptions. Using these methods, I found evidence of nonlinearities in the ability–earnings relationships that varied across levels of schooling. Specifically, I found strong evidence of diminishing returns to ability for those with 12 or fewer years of schooling and weak evidence of increasing returns to ability for those with at least some form of college education. Use of these methods also revealed the problem of ability-sorting into higher education as discussed in Cawley *et al.* (2000) and Heckman and Vytlacil (2001). As a general rule, the most able individuals go to college, while the least able do not, and thus problems necessarily arise when trying to estimate the return to schooling over the tails of the ability distribution. To combat this identification problem, I estimated the return to schooling over an ability support which is ‘common’ or ‘interior’ to both the college and no-college groups. I found that returns to schooling were indeed increasing for the *more* able over this region, as has been reported in previous work. Further, I find that the college wage premium has grown over time for individuals at all points in the ability distribution and that this growth may have been ‘smoother’ for those of higher cognitive ability.

Final Manuscript Received: October 2002

References

Ahn, H., and Powell, J. L. (1993). ‘Semiparametric estimation of censored selection models with a nonparametric selection mechanism’, *Journal of Econometrics*, Vol. 58, pp. 3–29.

Altonji, J. G. and Dunn, T. A. (1996). ‘Using siblings to estimate the effect of school quality on wages’, *Review of Economics and Statistics*, Vol. 78, pp. 665–71.

Angrist, J. and Krueger, A. B. (1991). ‘Does compulsory school attendance affect schooling and earnings?’, *Quarterly Journal of Economics*, Vol. 106, pp. 979–1014.

Ashenfelter, O. and Krueger, A. (1994). ‘Estimates of the economic return to schooling from a new sample of twins’, *American Economic Review*, Vol. 84, pp. 1157–73.

Ashenfelter, O. and Mooney, J. D. (1968). ‘Graduate education, ability, and earnings’, *Review of Economics and Statistics*, Vol. 50, pp. 78–86.

Ashenfelter, O. and Rouse, C. (1998). ‘Income, schooling and ability: evidence from a new sample of identical twins’, *Quarterly Journal of Economics*, Vol. 113, pp. 253–84.

Behrman, J. R., Rosenzweig, M. R. and Taubman, P. (1996). ‘College choice and wages: estimates using data on female twins’, *Review of Economics and Statistics*, Vol. 78, pp. 672–85.

Belman, D. and Heywood, J. S. (1991). ‘Sheepskin effects in the returns to education: an examination of women and minorities’, *Review of Economics and Statistics*, Vol. 73, pp. 720–24.

Blackburn, M. and Neumark, D. (1992). ‘Unobserved ability, efficiency wages, and interindustry wage differentials’, *Quarterly Journal of Economics*, Vol. 107, pp. 1421–36.

Blackburn, M. and Neumark, D. (1993). ‘Omitted-ability bias and the increase in the return to schooling’, *Journal of Labor Economics*, Vol. 11, pp. 521–44.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Blackburn, M. and Neumark, D. (1995). ‘Are OLS estimates of the return to schooling biased downward?’, *Review of Economics and Statistics*, Vol. 77, pp. 217–30.

Blundell, R. and Duncan, A. (1998). ‘Kernel regression in empirical microeconomics’, *Journal of Human Resources*, Vol. 33, pp. 62–87.

Card, D. (1995). ‘Using geographic variation in college proximity to estimate the return to schooling’, in Christofides L. N., Grant E. K. and Swidinsky R. (eds), *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*, University of Toronto Press, Toronto, pp. 201–22.

Card, D. (1999). ‘The causal effect of education on earnings’, in Ashenfelter O. C. and Card D. (eds), *Handbook of Labor Economics*, Vol. 3A, pp. 1801–63.

Card, D. (2001). ‘Estimating the return to schooling: progress on some persistent econometric problems’, *Econometrica*, Vol. 69, pp. 1127–60.

Card, D., and Krueger, A. B. (1992). ‘Does school quality matter? Returns to education and the characteristics of public schools in the United States’, *Journal of Political Economy*, Vol. 100, pp. 1–40.

Cawley, J., Conneely, K., Heckman, J. and Vytlacil, E. (1997). ‘Cognitive ability, wages, and meritocracy’, in Devlin B., Feinberg S. E., Resnick D. P. and Roeder K. (eds), *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*, Springer, New York, pp. 179–92.

Cawley, J., Heckman, J. and Vytlacil, E. (2000). ‘Three observations on wages and measured cognitive ability’, *Labour Economics*, Vol. 8, pp. 419–42.

Cawley, J., Heckman, J. and Vytlacil, E. (1999). ‘On policies to reward the value added by educators’, *Review of Economics and Statistics*, Vol. 81, pp. 720–28.

Conneely, K. and Uusitalo, R. (1997). ‘Estimating heterogeneous treatment effects in the becker schooling model’, Unpublished discussion paper, Industrial Relations Section, Princeton University.

DiNardo, J. and Tobias, J. L. (2001). ‘Nonparametric density and regression estimation’, *Journal of Economic Perspectives*, Vol. 15, pp. 11–28.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. Chapman & Hall, London.

Griliches, Z. (1977). ‘Estimating the returns to schooling: some econometric problems’, *Econometrica*, Vol. 45, pp. 1–22.

Griliches, Z. and Mason, W. M. (1972). ‘Education, income and ability’, *Journal of Political Economy*, Vol. 80, pp. S74–S103.

Grogger, J. and Eide, E. (1995). ‘Changes in college skills and the rise in the college wage premium’, *Journal of Human Resources*, Vol. 30, pp. 280–310.

Hansen, W. L., Weisbrod, B. A. and Scanlon, W. J. (1970). ‘Schooling and earnings of low achievers’, *American Economic Review*, Vol. 60, pp. 409–18.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Harmon, C., Oosterbeek, H. and Walker, I. (2003). ‘The returns to education: a review of evidence, issues and deficiencies in the literature’, *Journal of Economic Surveys*, (in press).

Harmon, C. and Walker, I. (1995). ‘Estimates of the economic return to schooling for the United Kingdom’, *American Economic Review*, Vol. 85, pp. 1278–86.

Harmon, C. and Walker, I. (1999). ‘The marginal and average returns to schooling in the UK’, *European Economic Review*, Vol. 43, pp. 879–87.

Heckman, J. J. (1976). ‘The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models’, *Annals of Economic and Social Measurement*, Vol. 5, pp. 475–92.

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998). ‘Characterizing selection bias using experimental data’, *Econometrica*, Vol. 66, pp. 1017–98.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Heckman, J., Lochner, L. and Todd, P. (2001). *'Fifty Years of Mincer Earnings Regressions'*, mimeo, University of Chicago, IL.

Heckman, J., Tobias, J. L. and Vytlacil, E. (2001). 'Four parameters of interest in the evaluation of social programs', *Southern Economic Journal*, Vol. 68, pp. 210–33.

Heckman, J., Tobias, J. L. and Vytlacil, E. (2003). 'Simple estimators for treatment parameters in a latent variable framework', *Review of Economics and Statistics*, (in press).

Heckman, J. and Vytlacil, E. (2000). 'The relationship between treatment parameters within a latent variable framework', *Economics Letters*, Vol. 66, pp. 33–39.

Heckman, J. and Vytlacil, E. (2001). 'Identifying the role of cognitive ability in explaining the level of and change in the return to schooling', *Review of Economics and Statistics*, Vol. 83, pp. 1–12.

Herrnstein, R. J. and Murray, C. (1994). *The Bell Curve*, Free Press, New York.

Hungerford, T. and Solon, G. (1987). 'Sheepskin effects in returns to education', *Review of Economics and Statistics*, Vol. 69, pp. 175–77.

Ichino, A. and Winter-Ebmer, R. (1999). 'Lower and upper bounds of returns to schooling: an exercise in IV estimation with different instruments', *European Economic Review*, Vol. 43, pp. 889–901.

Jaeger, D. and Page, M. (1996). 'Degrees matter: new evidence on sheepskin effects in returns to education', *Review of Economics and Statistics*, Vol. 78, pp. 733–40.

Meghir, C. and Palme, M. (1999). 'Assessing the effect of schooling on earnings using a social experiment', *SSE/EFI Working Paper Series In Economics and Finance No. 313*, Stockholm School of Economics, Stockholm.

Mincer, J. (1974). *Schooling, Experience and Earnings*, Columbia University Press, New York.

Moretti, E. (2003). 'Estimating the social returns to higher education: evidence from longitudinal and cross-sectional data', *Journal of Econometrics*, (in press).

Murnane, R. J., Willett, J. B. and Levy, F. (1995). 'The growing importance of cognitive skills in wage determination', *Review of Economics and Statistics*, Vol. 77, pp. 251–66.

Robinson, P. (1988). 'Root-N consistent semiparametric estimation', *Econometrica*, Vol. 56, pp. 931–54.

Rouse, C. (1999). 'Further evidence of the economic return to schooling from a new sample of twins', *Economics of Education Review*, Vol. 18, pp. 1999

Sianesi, B. and Van Reenen, J. (2003). 'The returns to education – a review of the macroeconomic literature', *Journal of Economic Surveys*, (in press).

Taber, C. (2001). 'The rising college premium in the eighties: return to college or return to unobserved ability?', *Review of Economic Studies*, Vol. 68, pp. 665–91.

Tobias, J. L. and Li, M. (2003). 'A finite sample hierarchical analysis of wage variation across public high schools: evidence from the NLSY and high school and beyond', *Journal of Applied Econometrics*, (in press).

Weisbrod, B. A. and Karpoff, P. (1968). 'Monetary returns to college education, student ability and college quality', *Review of Economics and Statistics*, Vol. 50, pp. 491–97.

Willis, R. J. and Rosen, S. (1979). 'Education and self-selection', *Journal of Political Economy*, Vol. 87, pp. S7–S36.

Yatchew, A. (1997). 'An elementary estimator of the partial linear model', *Economics Letters*, Vol. 57, pp. 135–43.

Yatchew, A. (1998). 'Nonparametric regression techniques in economics', *Journal of Economic Literature*, Vol. 36, pp. 669–721.

Yatchew, A. (1999). 'Differencing methods in nonparametric regression: simple techniques for the applied econometrician.', Unpublished Manuscript, University of Toronto.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix

We briefly describe our procedure for integrating out the effect of the characteristics **X** when presenting plots of the average treatment effect (ATE) over the ability support. Formally, we note:

$$\begin{aligned}\mathrm{ATE}(A_0) &\equiv E(y_1 - y_0 \mid A = A_0) = E_{X|A=A_0}[E(y_1 - y_0 \mid X, A = A_0)] \\ &\approx \sum_{i=1}^{n} E(y_1 - y_0 \mid X_i, A = A_0) w\left(\frac{A_i - A_0}{h_n}\right) \\ &= \sum_{i=1}^{n} X_i(\beta_1 - \beta_0) w\left(\frac{A_i - A_0}{h_n}\right) + m_1(A_0) - m_0(A_0).\end{aligned}$$

In the above, w is a normalized kernel weight, $w(r_i) = K(r_i)/\Sigma_j K(r_j)$, and the kernel function K is taken as a standard normal density function. To see why this procedure provides a consistent estimate of ATE(A_0), and the approximation listed above is a reasonable one, first suppose that A is discrete-valued.[27] In this case, we could estimate ATE(A_0) by simply averaging the ATE(X_i, $A = A_0$) values over those X_is where $A_i = A_0$. This is identical to choosing the weight function w to be an indicator function that equals one if $A_i = A_0$, and otherwise equals zero. As our ability variable A is continuous, we cannot implement such an approach. However, we can follow a similar logic and weight the treatment effects ATE($\mathbf{X}_i$, $A = A_0$) in an analogous manner according to the 'distance' from the A_i value to A_0. In large samples we will be able to choose the weighting function w so that only the ATE($\mathbf{X}_i$, A_i) values with A_i very close to A_0 are assigned significant weight in our estimator. Also note that setting $w(\cdot) = 1/n$ implies that our estimate of ATE(A_0) reduces to $\bar{\mathbf{X}}(\beta_1 - \beta_0) + m_1(A_0) - m_0(A_0)$. In this case, our estimator reduces to the often-used procedure of fixing the remaining explanatory variables to sample means. By choosing smaller values of h_n, we capture correlation between **X** and A and will, under certain regularity conditions, obtain consistent estimates of ATE(A_0).[28]

[27]Also note that we could estimate ATE(A_0) using Robinson's (1988) estimator, as $E(y_1 - y_0 \mid A) = E(X \mid A)(\beta_1 - \beta_0) + m_1(A) - m_0(A)$, and $E(X \mid A)$ can be estimated nonparametrically for each element of X.

[28]Formally, as our smoothing parameter h_n used in our kernel estimator approaches infinity, $w_i \rightarrow 1/n$ and thus the resulting estimator will consistently estimate E_X[ATE($\mathbf{X}$, $A = A_0$)]. This equally-weighted estimator integrates out the influence of **X** by integrating over the marginal distribution of **X**, while ATE(A_0) = $E(y_1 - y_0 \mid A = A_0)$ should be obtained by integrating over the conditional distribution of $X \mid A = A_0$, as described above. As a practical matter, however, we found that the resulting estimates using the above method are very similar to those obtained after fixing x values at sample means.

© Blackwell Publishing Ltd 2003

14680084, 2003, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1468-0084.00038 by Petr ala - <Shibboleth>-member@cuni.cz , Wiley Online Library on [28/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License