Measuring the Returns to Education Nonparametrically

Shawn W. Ulrick

November 8, 2005

Abstract

This article uses a nonparametric model of earnings to measure the returns to education. Under very general smoothness conditions, a nonparametric estimator reveals the true shape of the earnings profiles up to random sampling error. Thus, the nonparametric model should provide better predictions than its parametric counterpart. We find that the nonparametric model predicts very different estimated returns than standard Mincer formulations. Depending on the experience and education level, returns measured in log earnings estimated from nonparametric model can be nearly twice those obtained from the Mincer model. Finally, this paper examines what structural features parametric models should include.

1 U.S. Federal Trade Commission, 600 Pennsylvania Ave., NW, Washington, DC 20580. sulrick@ftc.gov; (202) 326-2327, fax: (202) 326-2625. I wish to thank Joel Horowitz, George Neumann, and John Geweke. I also wish to thank Qi Li for a helpful discussion about bandwidth selection in a large sample. The views in this article are those of the author and do not necessarily reflect those of the Federal Trade Commission or any individual Commissioner.
I. Introduction

Policy makers have long been interested in measuring the returns to education. One of the most widely used empirical models in this pursuit has been the Mincer (1974) formulation, in which log earnings is a quadratic in experience and linear in the other variables. Hundreds of studies have used this specification or slight variants of it. However, despite the widespread confidence in the Mincer model, simple hypothesis tests can show that it is misspecified. There is likely an interaction between experience and education, and experience is likely a higher order polynomial. This is perhaps alarming, since using a misspecified model may lead to incorrect measurements. Therefore, this paper uses a nonparametric model to estimate the returns to education. The nonparametric model assumes only very general smoothness conditions about the earnings function, and thus the estimator reveals the true shape of the earnings profiles up to random sampling error. It is thus less likely to give improper measurements of the returns to education.

We find that the nonparametric approach reveals information about the structure of the returns that the Mincer or similar parametric models are unable to show. We find that the returns vary by experience level (standard parametric approaches impose structure that assumes this feature away). Specifically, the returns to education appear greater for younger workers than for older ones. This pattern is especially noticeable when examining the returns to a college degree; the model predicts that the log earnings benefit to workers with 5 years of experience, for example, is nearly twice as high as for those with 35 years. We also find that the nonparametric predictions can be (at times
radically) different than those obtained using common parametric models. For younger workers (those freshly out of college), the nonparametrically estimated returns are almost twice as high as those obtained from common parametric approaches. This is an important result, since returns of younger workers are perhaps most relevant when formulating many policies such as education subsidies (those entering college would likely have expected returns closest to the recent graduates); the common parametric forms would greatly understate the private returns to this group.

There are important downsides to nonparametric models. First, nonparametric estimators suffer from “the curse of dimensionality.” That is, the variance of the estimated nonparametric model (at a given sample size) quickly increases as the number of variables (the dimension) increases; it often only takes a few variables before the model is rendered essentially useless. Second, even for relatively small dimensions, nonparametric estimators require large samples. However, if there is a parametric model that gives similar results as the nonparametric one in small dimensional problems, then it suggests that the parametric model may be sufficient in models having several variables or where limited data is available. Hence, this paper also examines what parametric models capture the shape of the nonparametric profiles.

The rest of this paper is organized as follows. Section II discusses important endogeneity issues in measuring the returns to education. The section summarizes recent surveys suggesting that (except in rare cases) using an Instrumental Variables approach may lead to greater biases than uncorrected estimates. Section III introduces the nonparametric model. Section IV discusses the nonparametric estimates. Section V
presents a few simple parametric models that appear to capture the shape of the nonparametric profiles and give similar estimated returns. Finally, Section VI concludes.

II. Endogeneity issues in measuring the returns to education

The literature on measuring the returns to education is very extensive, and the topics far exceed what one could hope to cover in a short overview. For in depth surveys, see Card (1995 and 1999) and Angrist and Krueger (1999). It is noteworthy that the majority of the research has discussed what is known as the “omitted ability bias.”

Suppose, for example, that earnings follow the Mincer formula,

\[ y_i = a + e_i b + \exp_i c + \exp_i^2 d + u_i, \]

where \( e_i \) is education, \( \exp \) is experience, and \( u_i \) is a disturbance. The concern is that \( u_i \) may be broken into ability \( a_i \) and other unmeasured factors \( e_i \). That is,

\[ u_i = a_i + e_i. \]

Those with high ability are more likely obtain higher levels of education, and thus \( u_i \) and \( e_i \) are likely correlated, violating a basic assumption of OLS. The common belief is that those with higher ability are likely to earn more, and thus the returns to education are likely to be overstated by OLS. Several instrumental variables (IV) approaches have been implemented.\(^2\) Methods of applying instruments to nonparametric estimators are considerably more involved and are only recently being developed (see, e.g., Hall and Horowitz, 2004; Blundel and Powell, 2003; Newey and Powell, 2003). An instrumental variables approach could be applied here, but recent evidence (summarized in the Card surveys) suggests that controlling for the omitted ability bias with instruments may induce more harm than good (except in rare cases).

Specifically, Card notes that many instrumental variables techniques tend to predict even higher returns to education than the corresponding OLS estimates – a counter intuitive result. He suggests that poor instruments cause many of the IV estimates to have a greater upward bias than the corresponding OLS ones. For example, he points out that the mandatory schooling laws that are the foundation of the influential Angrist and Kreuger (1991) study raise the education obtainment of individuals who otherwise would choose a low level of education. As Card discusses, if these people have higher than average marginal returns to schooling, the IV estimator would overstate the returns for the population as a whole.\(^3\)

Moreover, Card suggests that the omitted variables bias is actually quite small. He points to studies involving pairs of identical twins. If ability is the same for each member of a pair, standard panel data techniques lead to consistent estimates. Specifically, the ability component can be differenced out and thus the correlation between the education variable and the disturbance removed. In these studies, the OLS estimates of the returns are typically only about 10% larger than those obtained by applying panel data procedures.

For these reasons, this paper does not attempt to control for omitted ability with instruments. Thus, the returns estimated herein are probably slightly overstated. However, the bias is probably less than if we attempted an IV approach. Moreover, in applications, returns are often reported without correcting for the omitted ability bias (See, for example, Maxwell, 1999; Psacharopoulos, 1992, among others).

\(^3\) Berhman (1996) also writes about the lack of appropriate instruments. Additionally, see discussion on using poor instruments in the influential paper by Bound, et al. (1993) suggesting that “the cure [IV estimation] may be worse than the disease.”
III. Data

The data in this paper consists of observations on an individual’s race, age, years of education, and weekly earnings, from the March 1999 CPS. Experience is measured by \((age – education – 5)\). There is considerable research suggesting that the structure of wages has varied over time.\(^4\) As Murphy and Welch (1992) note, the changes are so large, it seems wrong to ignore them in an analysis on the structure of wages. Thus, rather than pooling several years of the CPS, the present paper focuses on one year.

The dataset is screened in the following ways. First, only observations on blacks and whites are included. Females are excluded from the study since their experience profiles are considerably different than those of males. For this reason, part time workers are excluded as well.\(^5\) Only individuals 18 years and older with 8 or more years of education and positive earnings are included. Individuals with experience less than one or greater than 40 are removed from the analysis, leaving only those in their prime working years. Summary statistics appear in Table 1.\(^6\)

The relevant year of the CPS measures education past high school according to degree obtained rather than by actual years of education. Some authors have converted (or suggested converting) the CPS measure to years of education according to the number of years it typically takes to complete a degree.\(^7\) For comparison to the literature, most models in this paper use the conversion. A mapping is provided in Table 2.\(^8\) The

\(^4\) See, for example, Bound and Johnson (1992), Katz and Autor (1999), Katz and Murphy (1992), Blackburn et al. (1990).

\(^5\) Part-time is defined as less than 35 h/week and 48 weeks/year.

\(^6\) The data is available from the author upon request.

\(^7\) See, for example, Card (1999) and Angrist and Krueger (1999). The scheme in this paper is also similar to one recommended by Jaeger (1997), with the largest difference being that we distinguish between individuals with/without a H.S. diploma, while Jaeger does not.

\(^8\) The Feb. 1990 CPS provides information on both degree obtainment and years of education. The conversion in table II is based off the mean years of education for each education level completed.
conversion is a reasonable one. There is evidence in the literature that degree is possibly of greater importance in determining wages than actual years of education.\(^9\) It is unlikely, for example, that an individual who completes a college degree after 17 years should earn more than one who completes the degree after 16 years, and thus those who took either amount of time to complete the degree should be treated as the same.

At this point it should be mentioned that the variables considered in this study are not exhaustive of those used in the literature. The nonparametric estimator used in this analysis suffers from the curse of dimensionality. Including a large number of variables will make the nonparametrically estimated earnings profiles too noisy to be of any use. For this reason, (aside from race) this paper includes as a control only experience, which is perhaps the most important and appears in almost any discussion on the issue. (The parametric models that follow in Section VI can be supplemented with additional variables added in linearly, if they are not interacted with education or experience. This is perhaps a more reasonable assumption than that education and experience are not interacted.)

**IV. A nonparametric model of earnings**

Let \(Z_i\) be an indicator variable denoting whether an individual is black. Let \(ed_i\) and \(exp_i\) be the years of education and experience, respectively, of individual \(i\), and let \(y_i\) be log weekly earnings. Then, earnings may be written as

\[
y_i = f(Z_i, ed_i, exp_i) + u_i,
\]

where \(u_i\) is a disturbance, and \(f\) is unknown. We estimate \(f\) nonparametrically.

---

\(^9\) See, for example, Jaeger and Page (1996). They discover that the sheepskin effect accounts for more than 50% of the earnings premium of a college degree over a high school diploma.
In implementing the estimator of the earnings function, education and experience can be treated as continuous.\textsuperscript{10} However, race is a categorical variable, and thus the traditional nonparametric kernel estimator would only work if separate regressions were implemented for blacks and whites. This would be an acceptable procedure, but a nonparametric estimator recently developed by Racine and Li (2004) will be used instead. This estimator smooths across the categorical data and will include observations of both races in estimating the earnings function of either race, resulting in an efficiency gain, of course at the expense of finite sample bias. An overview of the estimator and this paper’s method of bandwidth selection is presented in the appendix.

V. Nonparametric estimates and implications for the returns to education

Past analyses based on parametric models have imposed artificial structure on the earnings profiles. Many thus have failed to reveal certain characteristics visible with the nonparametric approach. We now discuss interesting features of the nonparametrically estimated profiles.

The nonparametrically estimated experience/earnings profiles are presented in Figure 1.\textsuperscript{11} Predicted earnings profiles (lines with circles) and 95% point wise confidence bands (solid lines) are displayed for black and white men having ten, twelve, sixteen, and eighteen years of education. Additionally, to provide an aid in visually determining fit, actual cell means (pluses) are plotted in the profiles. The actual cell means are obtained by averaging the log earnings of otherwise observationally identical individuals – that is

\textsuperscript{10} The nonparametric estimator smoothes over the education categories, and its predictions near education levels for which there are observations should be fairly sound. In the analysis that follows, the nonparametric model is never evaluated at points where there is no data.

\textsuperscript{11} The bandwidths were selected via the method discussed in the appendix. They were, for the education variable, .309; for the experience variable, 1.176; for the categorical race indicator, .028.
individuals of the same race, education, and experience.\textsuperscript{12} As can be seen, the nonparametrically estimated profiles capture the overall shape of the cell means, but with considerably less noise. Figure 2 displays the nonparametrically estimated education/earnings profiles (lines with circles) and 95\% point wise confidence bands (solid lines) for black and whites with five, fifteen, twenty-five, and thirty-five years of experience.

As can be seen, the experience earnings profiles have the familiar convex shape in which log-earnings increase rapidly for younger workers, flatten out, and eventually decrease. However, as will be discussed in the next section, the exact shape is considerably different than that of the Mincer model or slight variants of it. Figure 2 reveals that the earnings/education profiles are more or less linear in education, except at very high and low levels. This result is consistent with previous studies that have concentrated on examining whether the profiles are linear. However, as will be discussed in more detail below, there are differences in the slopes of the profiles amongst the experience levels, which standard Mincer-based parametric models cannot reveal.

Another interesting feature of the figures is the greater variance in expected earnings of the older workers (those with 25 or 35 years of experience), relative to the young (this is more noticeable in Figure 2). Some of the relatively high variance in the means for those with 35 years of experience can be explained by this group having fewer members than do the other brackets. But this explanation does not extend to those with twenty-five years of experience. Roughly the same number of individuals have close to

\textsuperscript{12} The cell means are actually an unbiased nonparametric estimator of the profiles. However, they have an enormous variance. For example, many of the cell means for the blacks have 95\% confidence intervals that are 2 units in length.
fifteen and twenty-five years of experience, and this number is nearly 20% higher than the number of individuals having close to five. For these groups the larger variance in expected earnings must be due to a greater variance in the underlying data. Perhaps the relatively low variance and more distinct linear pattern amongst younger earners is because their earnings are based more on the education signal since, for them, other gauges of productivity are unavailable. Pay for older workers may be more variable since it is based more on their experiences and past performance. Finally, another noticeable feature is that the variance in the expected earnings of blacks is greater than that of whites. This is not surprising, since the sample is only 8% black.

Returns to education can be measured by calculating the change in expected earnings for different levels of education. The year-to-year changes are too noisy to be very meaningful, but the average annual return over several years can be measured with more accuracy, for whites. For example, the difference in expected earnings between those with a high school and college degree can be computed. The average yearly return is obtained by dividing this difference by four. This figure has a smaller variance than the estimated year-to-year changes. In many applications, the average annual return to a college degree is more interesting than the returns of each specific year between high school and college, and thus the inability to accurately measure the latter is not always a disadvantage. Unfortunately, for the black group, even the average return over several years is difficult to measure with any accuracy, since their estimated profiles are so noisy.

---

13 The nonparametric estimator smoothes across experience cells. Hence, the number of observations at or near a particular experience level affect the variance of the estimates. Thus, these statements are based off of a neighborhood about a particular experience level. Specifically, the neighborhood includes the experience cell in question and the next highest and lowest cell.
(nonetheless, it is useful to include the blacks in the dataset for the efficiency gain in measuring the whites’ profiles).

The average yearly returns for whites were computed for three interesting changes in education: ten to twelve years, twelve to sixteen years, and sixteen to eighteen years. The results are displayed in Figure 3. The figure has one plot for each group. The pluses represent the average yearly change in expected log earnings for an additional year of education, as predicted by the nonparametric model. The horizontal bars represent 95% confidence intervals for the predicted change. For comparison, each plot also displays the predicted returns from parametric models introduced later in this paper. The predictions from these other models are represented by the symbols I – V. See the legend for details of the plotting symbols.

The most interesting feature of figure 3 is the average return to a year of education between high school and college: the average annual return varies considerably by experience level. Younger workers receive a greater return to a college degree than do older workers. For example, the average annual return to an individual with five years of experience is nearly 17%, while this figure is a much smaller 10% for those who have been working 35 years. Similarly, it appears that the return to a masters degree (i.e., a change in education of 16 to 18 years) is larger for younger workers than for older ones (although there is a large variance in the prediction for 35 years of experience). It has been well-documented that, since the 1970s, there has been an increase in real earnings for college graduates and a decrease for those who have only finished high school, thus raising the wage gap between the groups.\textsuperscript{14} Acs and Danziger (1993) provide discussion on the issue and a survey of articles investigating why this phenomenon is true. The
general consensus seems to be that there has been a shift from manufacturing based jobs to service oriented ones, which resulted in a lost of high paying jobs for high school educated people. The same technological changes that reduced the demand for low skilled workers have increased the demand for high-skilled ones. To the extent that this explanation is true, it seems plausible to expect that the phenomenon would have affected younger and older workers the same. However, the results of the nonparametric model suggest that, though the wage gap has indeed increased on average, the gap has increased disproportionately for younger workers. This fact reveals a need to study in more detail the events related to the increase in the wage gap. For example, do older college graduates not have the technical skills that have yielded an increase in the earnings of college graduates? Have older high school graduates gained skills during their work experiences relevant to the new structure of the economy? These are questions that research using parametric models has not sufficiently raised.

VI. Parametric

This paper has the advantage of a large dataset and includes only two variables beside education (i.e., experience and race). These facts make possible the use of the nonparametric estimator. Not all researchers have access to large datasets, and some researchers would prefer to include more control variables than does the present study. It is thus useful to consider what types of parametric models are flexible enough to capture the features of the profiles revealed by the nonparametric estimator; a parametric model that captures the features uncovered by the nonparametric estimator should provide a convenient summary of the data. This paper considers 5 parametric models (all estimated

---

14 See, for example, Katz and Autor (1999).
with the white group data). The first alternative is the Mincer (1974) specification of earnings; this is perhaps the most commonly used form in the literature:

\[ y_i = \alpha_0 + \beta ed_i + \gamma_1 exp_i + \gamma_2 exp_i^2 + u_i, \]

where \( \alpha_0, \beta, \) and \( \gamma \) are parameters, and \( u_i \) is an error term. The second model was chosen to be flexible while considering constraints on sample size. It is a very general six-degree polynomial expansion of education and experience that allows for complete interaction between the variables. The degree of the polynomial was chosen by using Akaike’s information criterion. This method weighs minimizing the sum of squared residuals against minimizing the number of parameters in a model.\(^{15}\)

\[ y_i = \sum_{j+k \leq 6} \gamma_{jk} ed_i^j \cdot exp_i^k + u_i. \]

Murphy and Welch (1992) suggest that the experience should be modeled as a quartic. Thus, the third parametric model is a slight variant of Mincer’s:

\[ y_i = \alpha_0 + \beta ed_i + 4 \sum_{j=1}^{4} \gamma_j exp_i^j + u_i. \]

The fourth model is a simplification of the very general Model (II). Specifically, Model (IV) is a fourth degree polynomial in experience, interacted with education. Education is linear, except for a vector of dummy variables (represented by \( d_i \)) indicating those with 9, 10, 16, 20 and 21 years of education.\(^{16}\) The fourth model is:

---

\(^{15}\) For details, see Greene (1997, page 787) and Geweke and Meese (1981). Other similar methods are also discussed in Geweke and Meese: Amemiya’s prediction criterion, Mallow’s \( C_p \) criterion, and Schwartz’s criterion. All methods suggested that model (II) should have a six-degree expansion, except Schwarz’s which suggested a four-degree expansion. The more general six degree was chosen.

\(^{16}\) Dummies for these education levels were chosen since the nonparametric education profiles in figure II suggest that the primary departure from linearity in the education variable occurs at very high and low levels of education, and since these represent years of education associated with degrees.
IV) \[ y_i = \alpha_0 + \beta ed_i + \sum_{j=1}^{4} \gamma_j exp_j + \sum_{j=1}^{4} \delta_j ed_i \cdot exp_j + d_i \phi + u_i, \]

where \( \delta_j \) and \( \phi \) are parameters. This model might be useful since it uses fewer parameters than (II), but is still fairly flexible and allows for the interaction. Finally, consider (V), which models education categorically and interacts education with experience. This specification is included to see if modeling education categorically improves the prediction.

V) \[ y_i = \sum_{k=1}^{s} \sum_{j=0}^{6} \eta_{jk} s_{kj} \cdot exp_j + u_i, \]

where \( s_{kj} \) represents indicator variables for each level of education represented in Table 2 (each education level -- beyond 8 years -- reported in the CPS), and \( \eta_{*} \) are parameters.

All five parametric models fail simple specification tests at any reasonable significance level, including the RESET and a basic nonparametric-based test.\(^{17}\) Despite this fact, (I) through (V) may still be useful in prediction if they capture the shape of the nonparametric profiles. Figure 4 makes this comparison for the Mincer model, and figure 5 does for Model (II). As can be seen, the Mincer model does a poor job of tracking, and thus we might not expect it to produce similar estimated returns (indeed, it does not, as discussed below). Model (II) follows quite closely (and, as will be discussed below, does a very good job of matching the nonparametric returns). Similar plots for the other models are excluded for brevity, but generally Models (IV and V) -- which include

\(^{17}\) In this test, the fitted values \( \hat{y}_i \) are obtained from the parametric regression. If the parametric model is properly specified a nonparametric mean regression of \( \hat{y}_i \) on \( y_i \) should result in a 45° line. Uniform confidence bands can be generated from such a regression, as well as a 45° line. If the model is properly specified, the 45° line should lie entirely inside the confidence bands. None of the regressions of \( \hat{y}_i \) on \( y_i \) lie within 90% uniform confidence intervals, and thus all of the parametric models are rejected.
interaction terms -- follow the nonparametric profiles more or less as well as (II); Model (III) -- which does not allow for interaction -- follows essentially no better than the Mincer.

The estimated returns from the various parametric specifications appear in Figure 3, along with the nonparametric estimates. As can be seen, Model (I) predicts approximately a 13.06% increase in earnings for each additional year of education. The model imposes the constraint that the returns are the same at every education and experience level and thus fails to capture the interaction made apparent by the nonparametric estimates. For certain experience levels, the predictions are off considerably. For example, the Mincer model understates the returns to a college degree by about 50%, for younger workers, and overstates the returns for older workers. The quartic Model (III) predicts an almost identical 12.93% return to each year of schooling (the predictions of Model (III) are excluded from the figure, since they would be indistinguishable from those of Model I), thus merely modeling experience as a quartic is not sufficient to address the shortfall of the Mincer model. These results could have important policy implications. Society may be less willing to subsidize education if it deems that the private returns are high enough that individuals enjoying the benefit of a degree should be borrowing instead. Those entering college would likely have expected returns closest to the recent graduates. The Mincer model would understate the returns of the recent graduates (because it averages across younger and older workers) and thus basing such a policy decision on the Mincer specification may be imprudent.

As can be seen, Model (II) -- which interacts education and experience -- does a much better job of mimicking the nonparametric predictions. Models (IV) and (V),
which include dummy variables for certain education levels, also follow the nonparametric predictions very well. In fact, Model (V) is usually almost identical to the nonparametric predictions. These result suggest that allowing for nonlinearities in education (e.g., the literature on sheepskin effects) may improve model fit, but allowing for interaction between education and experience is essential.

VII. Concluding remarks

This paper used a nonparametric model of earnings to measure the returns to education. This paper revealed structure to the returns that previous parametric approaches assumed away. It revealed that the return to a college degree is almost double for younger workers than for older ones. The returns to a masters degree is also higher for younger workers. The nonparametrically estimated returns were also considerably different than those obtained with the Mincer model. Depending on the experience and education level, the nonparametrically estimated returns could be almost double those obtained from it. Finally, the paper evidence suggesting that parametric models which interact experience and education may sufficiently well predict the returns.
Table 1: Summary statistics

<table>
<thead>
<tr>
<th>variable</th>
<th>description</th>
<th>sample means</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>blacks</td>
<td>whites</td>
</tr>
<tr>
<td>w</td>
<td>weekly earnings</td>
<td>607.5</td>
<td>833.5</td>
</tr>
<tr>
<td>y</td>
<td>log weekly earnings</td>
<td>6.162</td>
<td>6.403</td>
</tr>
<tr>
<td>ed</td>
<td>education</td>
<td>12.71</td>
<td>13.35</td>
</tr>
<tr>
<td>exp</td>
<td>experience</td>
<td>18.58</td>
<td>18.75</td>
</tr>
<tr>
<td>Z</td>
<td>black indicator</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Sample size: 2182 24857 27039
Table 2: Education cells

<table>
<thead>
<tr>
<th>Education cell</th>
<th>Continuous measure</th>
<th>indicator variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>9th grade</td>
<td>9</td>
<td>$s_1$</td>
</tr>
<tr>
<td>10th grade</td>
<td>10</td>
<td>$s_2$</td>
</tr>
<tr>
<td>11th grade or 12th grade and no diploma</td>
<td>11</td>
<td>$s_3$</td>
</tr>
<tr>
<td>High school graduate - High school diploma</td>
<td>12</td>
<td>$s_4$</td>
</tr>
<tr>
<td>Associate's degree in college - occupation</td>
<td>13</td>
<td>$s_5$</td>
</tr>
<tr>
<td>Associate's degree in college - academic</td>
<td>14</td>
<td>$s_6$</td>
</tr>
<tr>
<td>Bachelor's degree (e.g., BA, BS, AB)</td>
<td>16</td>
<td>$s_7$</td>
</tr>
<tr>
<td>Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)</td>
<td>18</td>
<td>$s_8$</td>
</tr>
<tr>
<td>Professional school degree (e.g.: MD, DDS, DVM, LLB, JD)</td>
<td>20</td>
<td>$s_9$</td>
</tr>
<tr>
<td>Doctorate degree (e.g., PhD, EdD)</td>
<td>21</td>
<td>$s_{10}$</td>
</tr>
</tbody>
</table>
Figure 1: Nonparametrically estimated experience/earnings profiles

Note: circles are nonparametric profiles; pluses are actual cell means; solid lines are 95% point-wise confidence bands
Figure 2: Nonparametrically estimated education/earnings

Note: circles are nonparametric profiles; solid lines are 95% point-wise confidence bands
Figure 3: Returns to education

Note: Plotting symbols are as follows: + - nonparametric return; bars (“-“) - 95% point-wise confidence bands for the nonparametric prediction; I - return predicted Mincer model; II – return predicted by model (II) (experience polynomial degree six, interacted with education); IV - return predicted by model (IV) (experience polynomial degree four, interacted with education); V – return predicted by model (V) (categorical education); note: plotting for symbols for quartic model (III) not shown since they are indistinguishable from the Mincer models.
Figure 4: Earnings profiles for whites: Mincer (circles) vs. nonparametric (pluses)
Figure 5: Earnings profiles for whites: Model (II) (circles) vs. nonparametric (pluses)
Appendix

The purpose of this appendix is to introduce the nonparametric regression estimator presented in Racine and Li (2004) (R&L). The estimator allows a nonparametric regression with both continuous and categorical variables. The form of the estimator presented below is less general than in R&L, to simplify the discussion.

Let $Z_i$ be an indicator variable denoting that individual $i$ is black, as defined in the text. Let there be $n$ individuals. Let $ed_i$ and $exp_i$ be as defined in the text. Define $X_i = (ed_i, exp_i)$. For estimation purposes, $X_i$ may be treated as continuous. Write log annual earnings $Y_i$ as

$$Y_i = f(Z_i, X_i) + u_i,$$

where $u_i$ is a disturbance, and $f$ is unknown. The estimator presented in R&L uses a kernel method. A separate kernel is required for the continuous and discrete components of $(Z_i, X_i)$. Let $W(\cdot)$ denote the kernel corresponding to the continuous component $X$. In this paper, $W$ was chosen to be the standard normal density function. The kernel corresponding to the discrete component is defined as follows. Let

$$L(Z_i, z) = \begin{cases} 1 & \text{if } Z_i = z \\ \lambda & \text{if } Z_i \neq z \end{cases},$$

where $\lambda$ is a smoothing parameter. Thus, the kernel function corresponding to the discrete component takes on the value 1 if $Z_i = z$, and $\lambda$ otherwise. Let

$$K_i(z, x) = L(Z_i, z) h^2 W((X_i - x)/h),$$

where $h$ is the smoothing parameter. Thus, $K_i$ is the product of the kernels corresponding to the discrete and continuous components. Racine and Li show that $f$ may be estimated by

$$\hat{f}(z, x) = \frac{\sum_{i=1}^n Y_i K_i(z, x)}{\sum_{i=1}^n K_i(z, x)},$$

if $\lambda, h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$. Note how the estimator smoothes across the categorical variable $Z_i$. Observations of whites are weighted by $\lambda$ and included in estimating the earnings function of blacks, as are blacks in estimating the earnings function of whites. Note that for $\lambda = 0$, $\hat{f}$ amounts to implementing a separate nonparametric regression on blacks and whites, and, for $\lambda = 1$, $\hat{f}$ amounts to estimating the pooled earnings function.

R&L suggest a cross-validation approach to choosing the optimal bandwidths. The method is to choose $(\lambda, h)$ to minimize

$$CV = \sum_{i=1}^n [Y_i - \hat{f}_{-i}(X_i)]^2 M(X_i),$$

where $M$ is a weight function and
\[
\hat{f}_{-i}(X_i) = \frac{\sum_{j \neq i} Y_{j}K_{j}(Z_{i}, X_{i})}{\sum_{j \neq i} K_{j}(Z_{i}, X_{i})}.
\]

For the sample size in this paper \((n = 27,000)\), the method is not practical. Instead, the bandwidths were chosen by noting that, as shown in R&L, the optimal rate of the bandwidths \(\lambda\) and \(h\) are \(c_{l}n^{-2/6}\) and \(c_{h}n^{-1/6}\), respectively, for some constants \(c_{l}\) and \(c_{h}\). From the data, a subsample of size 1000 was taken and \(\lambda^*\) and \(h^*\) were chosen to minimize \(CV\). The constants \(c_{l}\) and \(c_{h}\) were chosen to solve \(\lambda^* = c_{l}1000^{-2/6}\) and \(h^* = c_{h}1000^{-1/6}\). This process was repeated several times, and the resulting \(c_{l}\)'s and \(c_{h}\)'s were averaged to obtain \(\bar{c}_{l}\) and \(\bar{c}_{h}\), respectively. The bandwidths used to estimate \(f\) in the full sample were chosen to be \(\overline{c}_{l}n^{-2/6}\) and \(\overline{c}_{h}n^{-1/6}\).
References


