Notes on Policies to Reward the Value Added by Educators

By John Cawley, James Heckman, and Edward Vytlacil
Review of Economics and Statistics (1999)

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 350, Winter 2021

I. Introduction

- The performance of teachers and schools is often measured by the standardized test scores of their students.
- For example, the Tennessee study of randomized variation in class size in the early school grades uses gains in percentiles in nationally standardized tests to measure the gains in performance from smaller class sizes.
- South Carolina and Tennessee have begun measuring the "value added" of educators by measuring the gain in student test scores.
- Public school administrators and teachers in Dallas, Texas, receive cash awards based on changes in average test scores.
- Test scores are interesting only to the extent that they predict some measure of adult achievement such as educational attainment or wages.
- This paper investigates the relationship between test scores and log wages.

- We use the National Longitudinal Survey of Youth (NLSY) to examine whether socioeconomic outcomes are a linear function of test scores.
- Log wages is our outcome of interest.
- The test scores we use are from the Numerical Operations and Math Knowledge subtests of the Armed Services Vocational Aptitude Battery (ASVAB).
- Using linear spline regressions, we find that the curvature of the return function of wages for test scores is nonlinear and varies dramatically depending on the test, the transformation of the test score, and the age at which wages were measured.
- Our findings imply that the average gain in test scores is an inadequate measure of school performance.

II. The Benefits of Linearity For Estimating Value Added

- Let y_i be the socioeconomic outcome measure of interest for individual i, where we assume that y_i is a scalar.

- Let $g(x) = E(y|x)$ where x is a vector of test scores.

- Let x_0^i be the test score for individual I before the program and $x_0^i + \Delta^i$ be the test score after the program, so that the test score gain for individual i is Δ^i.

- For example, the program may be a year of schooling at the high school being evaluated, so that we are examining the gain in test scores for a particular individual spending one year at the given high school.

- Then the expected gain in the outcome measure for individual i is $E\left(y\middle|x_0^i + \Delta^i\right) - E\left(y\middle|x_0^i\right) = g\left(x_0^i + \Delta^i\right) - g\left(x_0^i\right)$.

A. Linear Outcome Function

- If $g(x)$ is a linear function, we can represent it as $g(x) = \alpha + x\beta$, and we have

$$E(y|x_0 + \Delta) - E(y|x_0) = \Delta\beta. \tag{1}$$

- Let $F_{x_{0,\Delta}}$ denote the joint distribution of x_0 and Δ and F_Δ denote the marginal distribution of Δ.

- The average expected gain in the outcome measure is

$$\begin{aligned} \int (E(y|x_0 + \Delta) - E(y|x_0))\, dF_{x_0,\Delta} &= \int \Delta dF_\Delta \beta \\ &= \bar{\Delta}\beta. \end{aligned} \tag{2}$$

B. Nonlinear Outcome Function

- If the outcome function g(x) is not linear, then the average expected gain is

$$\int (E(y|x_0+\Delta) - E(y|x_0))\, dF_{x_{0,\Delta}}$$

$$= \int (g(x_0+\Delta) - g(x_0))\, dF_{x_{0,\Delta}} \qquad (3)$$

$$= \int g(x_0+\Delta) dF_{xsb0+\Delta} - \int g(x_0)\, dF_{x_0}$$

where $F_{x0+\Delta}$ is the distribution of test scores after the intervention.

- To determine value added, it is no longer sufficient to know the mean test score gain; the school with the highest average test score gain would not necessarily be that with the highest average socioeconomic outcome gain.

- Figure 1 is a graph of a nonlinear outcome function.
- This figure is based on empirical results that we discuss in the next section.
- It shows our linear spline estimate of the relationship at age 30 between log wages and the Numerical Operations percentile scores.
- The outcome function has a slope of 0.015 for scores less than 25, and a slope of 0.004 for scores between 25 and 75.
- Thus, a one-point increase in the scores for those in the bottom of the test score distribution has an effect on log wages that is more than three times as large as the same one-point increase for those in the middle of the test score distribution.

III. Data Set

- We use the National Longitudinal Survey of Youth (NLSY) for our empirical analysis.

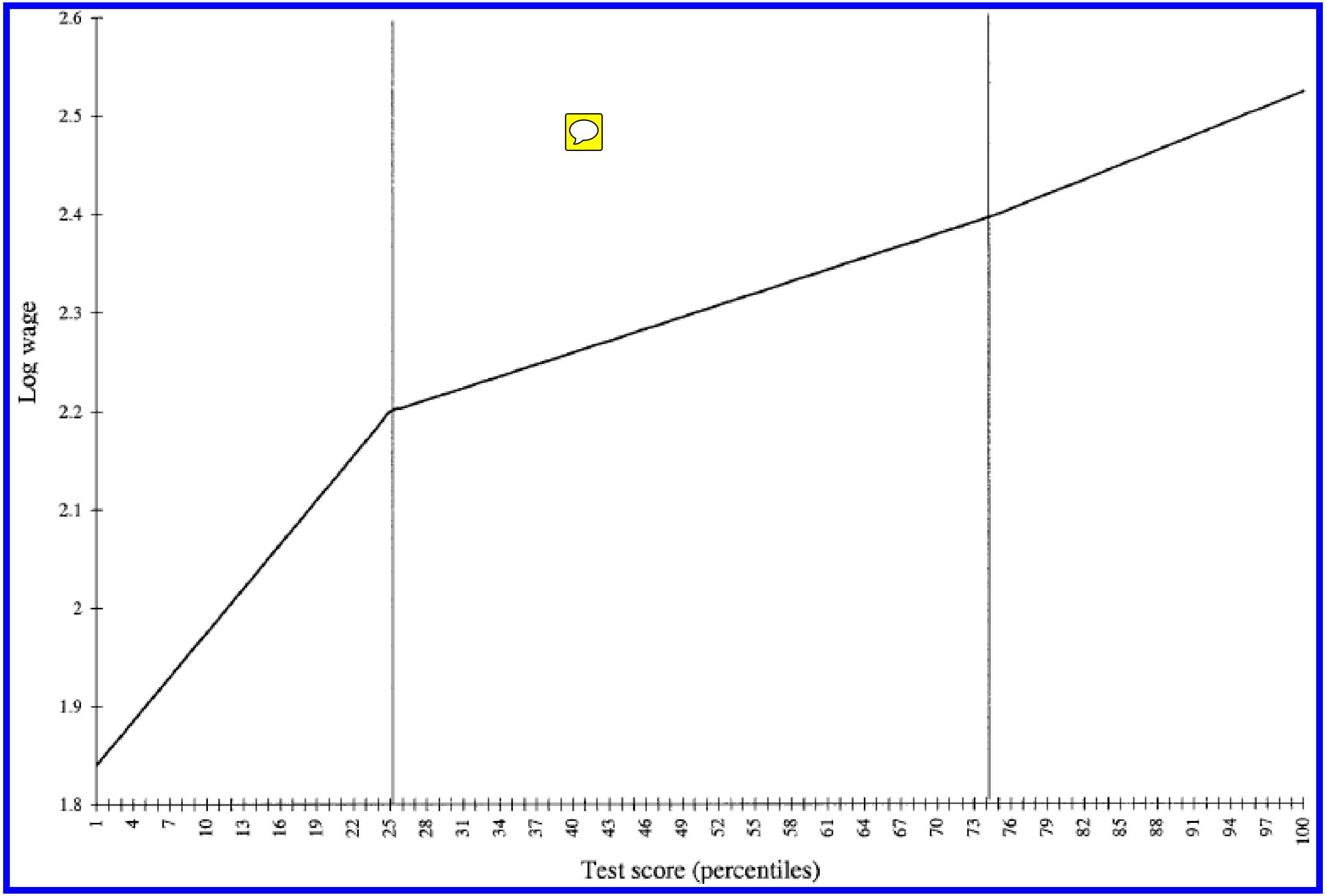
- The NLSY is designed to represent the entire population of American youth and consists of a randomly chosen sample of 6,111 U.S. civilian youths, a supplemental sample of 5,295 randomly chosen minority and economically disadvantaged civilian youths, and a sample of 1,280 youths on active duty in the military.

- All youths were between thirteen and twenty years of age in 1978 and were interviewed annually starting in 1979.

- For our analysis, we examine white men with valid test score data who are not currently enrolled in school and earn an hourly wage between $0.50 and $1,000 in 1990 dollars.

- (All results of this paper are reported in 1990 dollars.)6 Our resulting sample size is 3,528.

Figure 1: Nonlinear Output Function



IV. Analysis of Test Scores

- Figure 2 shows the sample densities of the Numerical Operations raw scores.

- The distribution is skewed to the left. We do not plot the density of the percentile scores, since it is uniform by construction.

- Figure 3 plots Numerical Operations raw scores versus Numerical Operations percentile scores. This figure indicates that the relationship between raw scores and percentile scores is approximately linear for raw scores between 21 and 1.5. While the relationship is linear within this interval, it is highly nonlinear overall.

- Figure 4 shows the smoothed sample density of the raw Math Knowledge scores. The distribution is left-skewed, although the left tail is small. Overall, the density is more uniform than the distribution of Numerical Operations scores.

- Figure 5 plots Math Knowledge raw scores versus percentile scores. Again, the relationship between raw scores and percentile scores is approximately linear for raw scores between 21 and 1.5, with the relationship nonlinear overall.

Figure 2: Density of Numerical Operations Raw Scores Kernel Density Estimate Using a Triangle Kernel and Silverman's Optimal Bandwidth

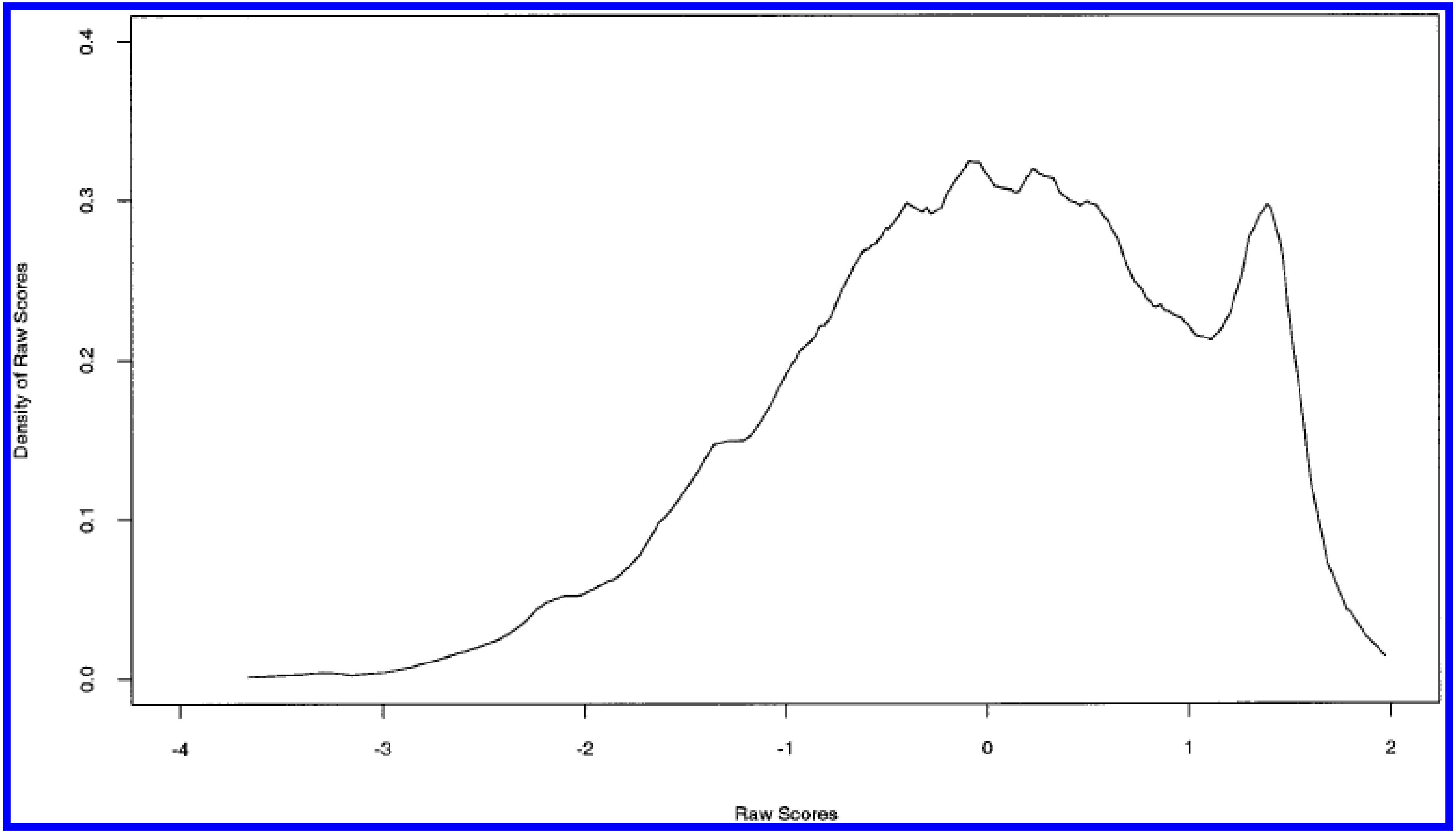
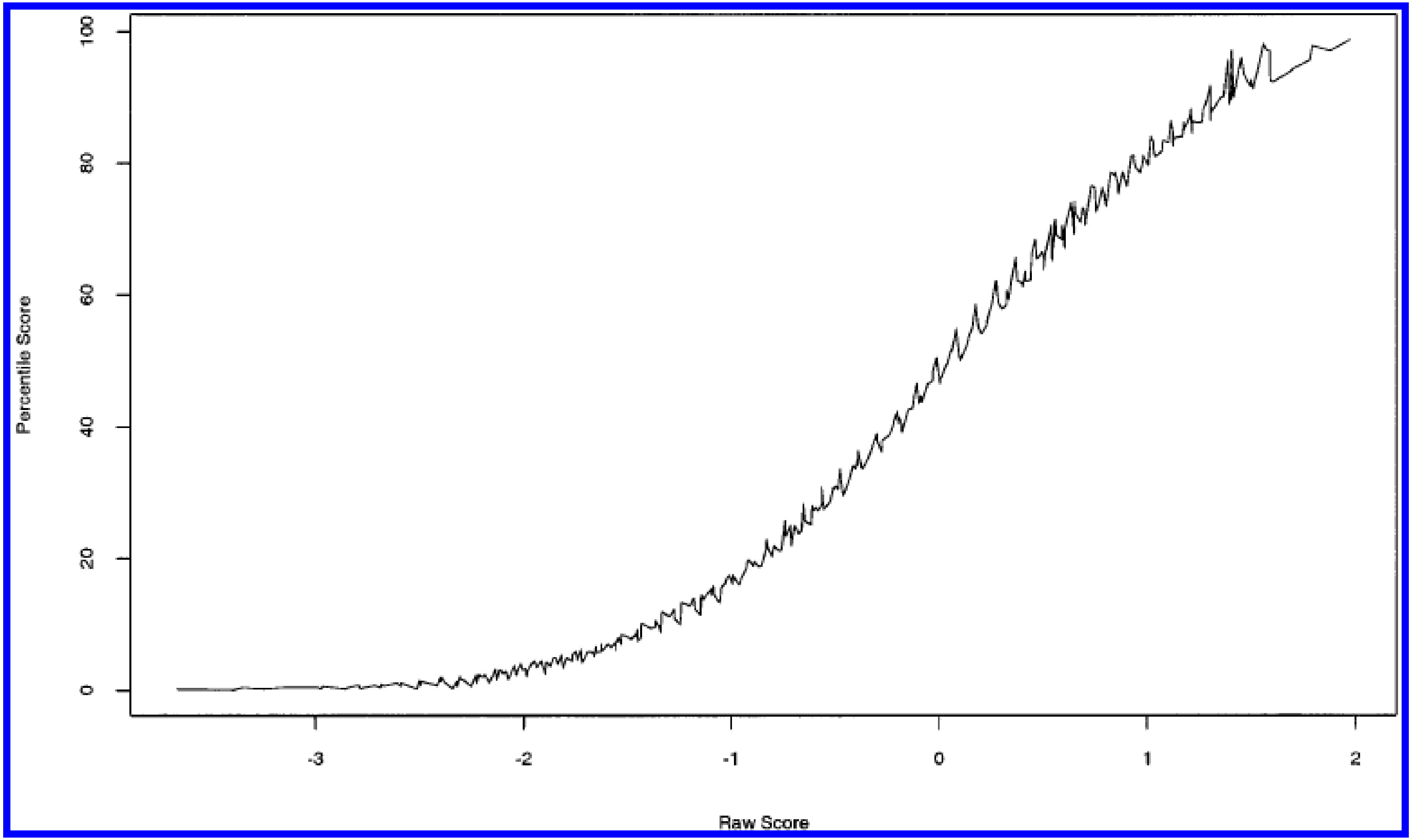
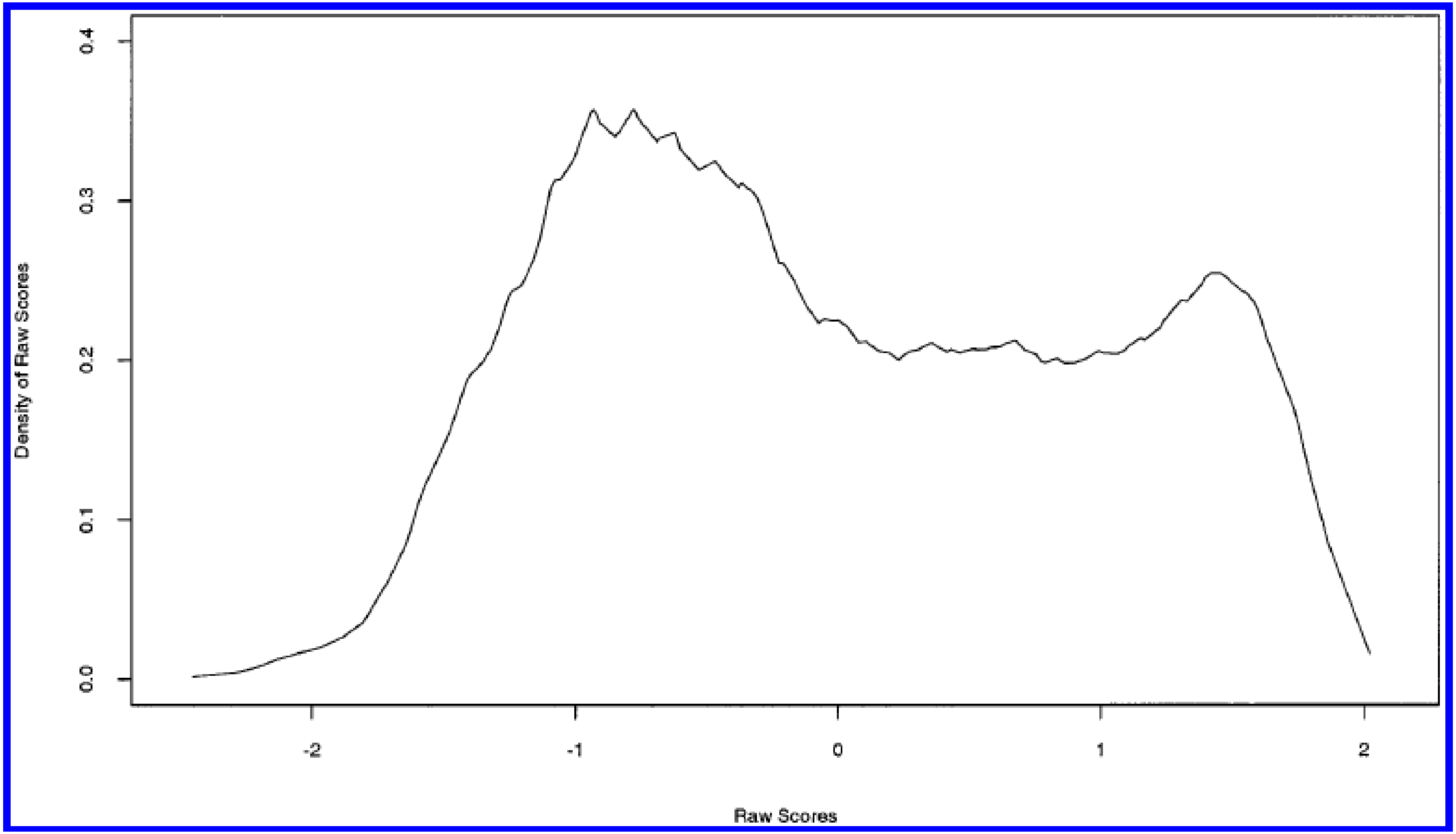


Figure 3: Numerical Operations Raw Score Versus Percentile Score



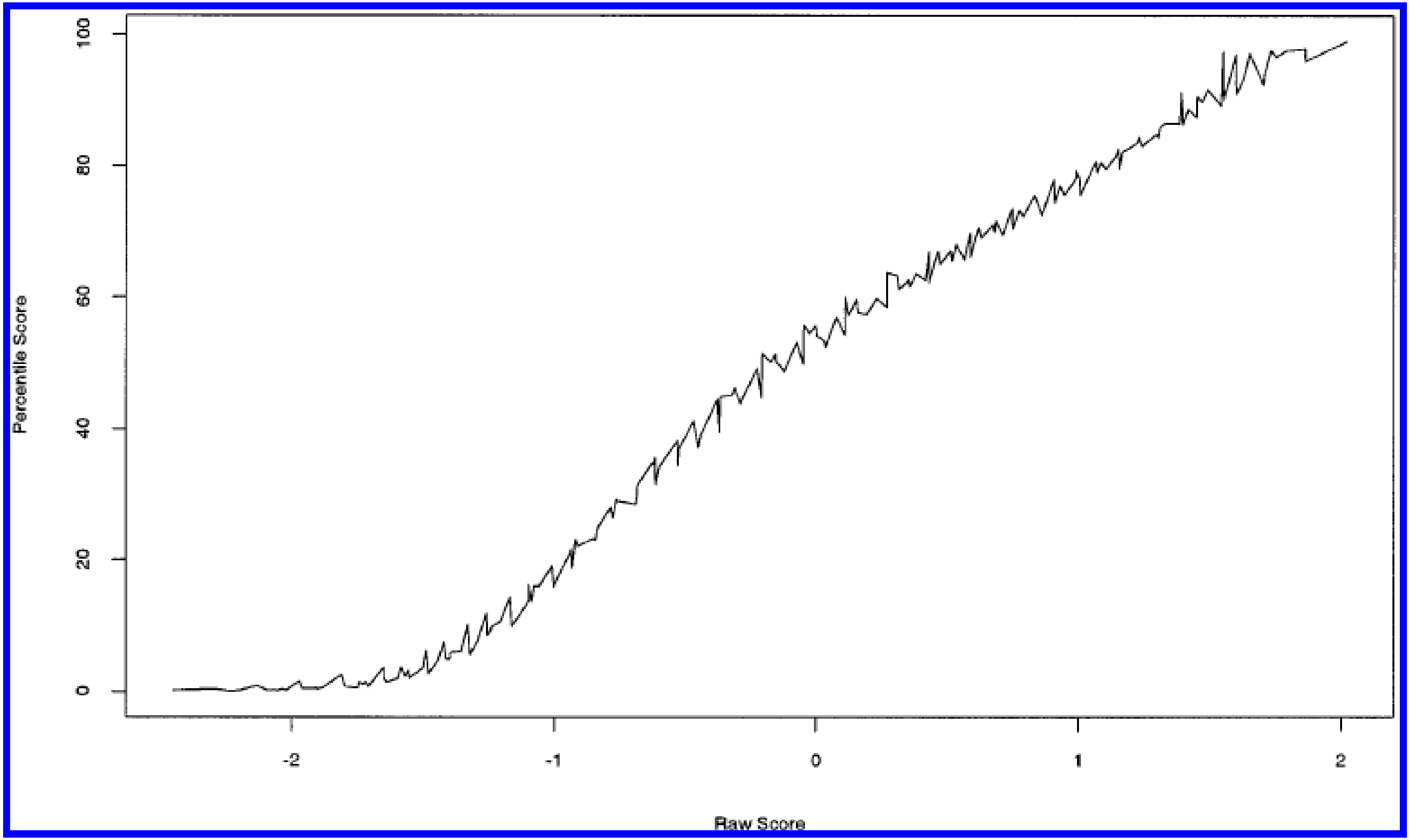
Note: Number of Observations: 3528

Figure 4: Density of Math Knowledge Raw Scores Kernel Density Estimate Using a Triangle Kernel and Silverman's Optimal Bandwidth



Note: Number of Observations: 3528

Figure 5: Math Knowledge Raw Score Versus Percentile Score



Note: Number of Observations: 3528

V. Splines

A. Numerical Operations Scores

- In table 1, we examine the relationship between log wages and Numerical Operations percentile (NOP) scores for wages measured at odd-numbered ages 19 through 31.

- (We delete the results for even numbers for the sake of brevity.)

- We regress the age-specific log wage on NOP scores and indicator variables for the year and for the local unemployment rate.

- We depart from the common practice of including years of schooling, work experience, and job tenure as regressors.

- (We use a reduced form to allow the effects of NOP to work through these omitted regressors.)

Table 1: Linear Splines Using Numerical Operations Percentile Scores

OLS REGRESSION WITH EICKER-WHITE STANDARD ERRORS
DEPENDENT VARIABLE: LOG HOURLY WAGES IN 1990 DOLLARS

Age 19	Variable	Coeff	Std Error	T-stat	P-val
	Num Op Percentiles	0.001	0.0012	1.14	0.255
	Num Op Per <25th Quantile	−0.002	0.0052	−0.37	0.715
	Num Op Per >75th Quantile	−0.003	0.0039	−0.7	0.484
	$R^2 = .11$, N = 638				
Age 21	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Percentiles	0.001	0.0009	0.62	0.534
	Num Op Per <25th Quantile	0.008	0.0031	2.43	0.015
	Num Op Per >75th Quantile	0.002	0.0031	0.56	0.578
	$R^2 = .11$, N = 1370				
Age 23	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Percentiles	0.001	0.0009	1.39	0.163
	Num Op Per <25th Quantile	0.006	0.0032	1.96	0.05
	Num Op Per >75th Quantile	0.004	0.0026	1.42	0.155
	$R^2 = .08$, N = 2021				
Age 25	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Percentiles	0.003	0.001	3.58	0
	Num Op Per <25th Quantile	0.008	0.0031	2.73	0.006
	Num Op Per >75th Quantile	−0.001	0.0027	−0.19	0.849
	$R^2 = .11$, N = 2321				

Table 1: Linear Splines Using Numerical Operations Percentile Scores

Age 27	Variable	Coeff	Std Error	T-stat	P-val
	Num Op Percentiles	0.003	0.0009	3.51	0
	Num Op Per <25th Quantile	0.009	0.0028	3.24	0.001
	Num Op Per >75th Quantile	0.002	0.0024	0.98	0.328
	R^2 = .13, N = 2258				
Age 29	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Percentiles	0.004	0.0009	4.15	0
	Num Op Per <25th Quantile	0.01	0.003	3.22	0.001
	Num Op Per >75th Quantile	0.003	0.0027	0.95	0.342
	R^2 = .14, N = 2085				
Age 31	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Percentiles	0.003	0.0011	2.99	0.003
	Num Op Per <25th Quantile	0.009	0.0045	2.04	0.041
	Num Op Per >75th Quantile	0.006	0.0028	1.95	0.051
	R^2 = .14, N = 1570				

1) Sample includes all valid employed out-of-school white males. NLSY sample weights are used.
2) Regressors are the Numerical Operations percentile score standardized by age and indicator variables for the year, the local unemployment rate, region of residence, and living in an urban area.

- In table 2, we repeat the analysis of table 1 for Numerical Operations raw (NOR) scores.
- For each age 21 or older, we find much larger slopes for the left tail, although the standard error is also consistently large.
- The nonlinearity in the left tail is significant at the 0.10 level only for age 21.
- The coefficient on the 75th-percentile knot is also sometimes large, although of different signs for different ages, consistently has a large standard error, and is statistically significant at the 0.10 level only for age 23.
- We conclude that, for some ages, there is substantial evidence of nonlinearity in the left tail of NOR scores; however, the curvature is poorly estimated.

B. Math Knowledge Scores

Table 2: Linear Splines Using Numerical Operations Raw Scores

OLS Regression with Eicker-White Standard Errors
Dependent Variable: Log Hourly Wages in 1990 Dollars

Age 19	Variable	Coeff	Std Error	T-stat	P-val
	Num Op Raw	0.074	0.0413	1.78	0.075
	Num Op Raw <25th Quantile	−0.11	0.1108	0.99	0.321
	Num Op Raw >75th Quantile	−0.134	0.0961	−1.39	0.164
	$R^2 = .11$, N = 638				
Age 21	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Raw	0.029	0.0301	0.95	0.343
	Num Op Raw <25th Quantile	0.11	0.0622	1.78	0.076
	Num Op Raw >75th Quantile	0.029	0.0858	0.34	0.734
	$R^2 = .11$, N = 1370				
Age 23	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Raw	0.032	0.027	1.19	0.233
	Num Op Raw <25th Quantile	0.1	0.0626	1.6	0.11
	Num Op Raw >75th Quantile	0.139	0.0745	1.87	0.062
	$R^2 = .08$, N = 2021				
Age 25	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Raw	0.116	0.0292	3.96	0
	Num Op Raw <25th Quantile	0.68	0.0584	1.16	0.248
	Num Op Raw >75th Quantile	−0.004	0.0725	−0.05	0.957
	$R^2 = .11$, N = 2321				

Table 2: Linear Splines Using Numerical Operations Raw Scores

Age 27	Variable	Coeff	Std Error	T-stat	P-val
	Num Op Raw	0.114	0.0262	4.35	0
	Num Op Raw <25th Quantile	0.084	0.0535	1.58	0.114
	Num Op Raw >75th Quantile	0.016	0.0663	0.24	0.808
	$R^2 = .13$, N = 2258				
Age 29	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Raw	0.126	0.0275	4.57	0
	Num Op Raw <25th Quantile	0.093	0.0571	1.63	0.103
	Num Op Raw >75th Quantile	0.05	0.0676	0.74	0.457
	$R^2 = .15$, N = 2085				
Age 31	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Num Op Raw	0.116	0.0361	3.22	0.001
	Num Op Raw <25th Quantile	0.114	0.0763	1.49	0.135
	Num Op Raw >75th Quantile	0.131	0.0867	1.51	0.131
	$R^2 = .15$, N = 1570				

1) Sample includes all valid employed out-of-school white males. NLSY sample weights are used.
2) Regressors are the Numerical Operations raw score standardized by age and indicator variables for the year, local unemployment rate, region of residence, and living in an urban area.

- Table 3 repeats the analysis of table 1, but for Math Knowledge percentile (MKP) scores.

- For age 31, we find substantial evidence of higher slopes (and thus nonlinearity) in both the left and right tails; the difference in slope between the middle and the tails is significant at the 0.05 level. A 1% increase in MKP score has a much larger effect for those with low or high scores.

- For ages 23 to 29, we find the same pattern of higher slopes in the left and right tails, but the difference in slopes is smaller and generally insignificant due to the large standard errors.

- We find very different results for the youngest ages: For ages 19 to 21, we find a lower slope in the right tail with the nonlinearity being significant at the 0.05 level for age 21.

Table 3: Linear Splines Using Mathematical Knowledge Percentile Scores

OLS Regression with Eicker-White Standard Errors
Dependent Variable: Log Hourly Wages in 1990 Dollars

Age 19	Variable	Coeff	Std Error	T-stat	P-val
	Math Know Percentiles	0.001	0.0011	1.07	0.284
	Math Know Per <25th Quantile	−0.008	0.0057	−1.34	0.18
	Math Know Per >75th Quantile	−0.005	0.004	−1.37	0.17
	$R^2 = .11$, N = 638				
Age 21	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Percentiles	0.001	0.0008	1.39	0.165
	Math Know Per <25th Quantile	0.003	0.0032	1.01	0.313
	Math Know Per >75th Quantile	−0.007	0.0031	−2.28	0.023
	$R^2 = .1$, N = 1370				
Age 23	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Percentiles	0.001	0.0008	1.5	0.133
	Math Know Per <25th Quantile	0.003	0.0034	1	0.317
	Math Know Per >75th Quantile	0.003	0.0027	0.97	0.333
	$R^2 = .07$, N = 2021				
Age 25	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Percentiles	0.001	0.0009	1.65	0.098
	Math Know Per <25th Quantile	0.01	0.0034	2.95	0.003
	Math Know Per >75th Quantile	0.007	0.0027	2.63	0.009
	$R^2 = .1$, N = 2321				

Table 3: Linear Splines Using Mathematical Knowledge Percentile Scores

Age 27	Variable	Coeff	Std Error	T-stat	P-val
	Math Know Percentiles	0.003	0.0008	3.18	0.001
	Math Know Per <25th Quantile	0.005	0.003	1.57	0.115
	Math Know Per >75th Quantile	0.005	0.0025	2.1	0.036
	$R^2 = .11$, N = 2258				
Age 29	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Percentiles	0.004	0.0008	5.13	0
	Math Know Per <25th Quantile	0.002	0.003	0.76	0.449
	Math Know Per >75th Quantile	0.006	0.0026	2.28	0.022
	$R^2 = .15$, N = 2085				
Age 31	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Percentiles	0.003	0.001	2.79	0.005
	Math Know Per <25th Quantile	0.009	0.0036	2.38	0.017
	Math Know Per >75th Quantile	0.01	0.0033	3.11	0.002
	$R^2 = .15$, N = 1570				

1) Sample includes all valid employed out-of-school white males. NLSY sample weights are used.
2) Regressors are the Mathematical Knowledge percentile score standardized by age and indicator variables for the year, local unemployment rate, region of residence, and living in an urban area.

- Table 4 presents similar results for Math Knowledge raw (MKR) scores.
- For the oldest age (31), we find substantial evidence of nonlinearity in both the left and right tails; the difference in slope between the middle and the tails is significant at the 0.10 level.
- For this age, a 1% increase in MKR has a much larger effect for those with low or high MKR scores.
- We generally find qualitatively similar but weaker results for ages 23 to 29, with the nonlinearity in both tails significant at the 0.10 level only for age 25.

C. Sensitivity Analysis

Table 4: Linear Splines Using Mathematical Knowledge Raw Scores

OLS Regression with Eicker-White Standard Errors
Dependent Variable: Log Hourly Wages in 1990 Dollars
Sample: White Males

Age 19	Variable	Coeff	Std Error	T-stat	P-val
	Math Know Raw	0.04	0.0425	0.95	0.343
	Math Know Raw <25th Quantile	−0.352	0.2052	−1.72	0.086
	Math Know Raw >75th Quantile	−0.095	0.1006	−0.94	0.347
	$R^2 = .12$, N = 629				
Age 21	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Raw	0.043	0.0279	1.52	0.128
	Math Know Raw <25th Quantile	0.058	0.0957	0.61	0.544
	Math Know Raw >75th Quantile	−0.181	0.0762	−2.38	0.017
	$R^2 = .13$, N = 1356				
Age 23	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Raw	−0.009	0.0278	−0.34	0.736
	Math Know Raw <25th Quantile	0.139	0.1067	1.3	0.194
	Math Know Raw >75th Quantile	0.136	0.071	1.92	0.055
	$R^2 = .11$, N = 2003				
Age 25	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Raw	0.011	0.0293	0.38	0.704
	Math Know Raw <25th Quantile	0.226	0.1047	2.16	0.031
	Math Know Raw >75th Quantile	0.132	0.0731	1.81	0.071
	$R^2 = .15$, N = 2294				

Table 4: Linear Splines Using Mathematical Knowledge Raw Scores

Age 27	Variable	Coeff	Std Error	T-stat	P-val
	Math Know Raw	0.017	0.0265	0.64	0.521
	Math Know Raw <25th Quantile	0.142	0.081	1.75	0.08
	Math Know Raw >75th Quantile	0.101	0.0665	1.52	0.128
	$R^2 = .17$, N = 2236				
Age 29	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Raw	0.066	0.0277	2.36	0.018
	Math Know Raw <25th Quantile	0.003	0.0821	0.04	0.971
	Math Know Raw >75th Quantile	0.086	0.0662	1.29	0.195
	$R^2 = .19$, N = 2055				
Age 31	**Variable**	**Coeff**	**Std Error**	**T-stat**	**P-val**
	Math Know Raw	0.026	0.0357	0.74	0.46
	Math Know Raw <25th Quantile	0.257	0.0942	2.73	0.006
	Math Know Raw >75th Quantile	0.182	0.0884	2.06	0.04
	$R^2 = .19$, N = 1543				

1) Sample includes all valid employed out-of-school white males. NLSY sample weights are used.
2) Regressors are the Mathematical Knowledge raw score standardized by age and indicator variables for the year, local unemployment rate, region of residence, and living in an urban area.

VI. Conclusion