

ECON 2320 – Economics of Labor and Population

Problem Set 1

Samuel Brown

Fall 2011

Due: October 03

1. (a) A scatterplot of log wages versus education, and the corresponding bivariate regression*, can be seen in Figure 1. The result of this bivariate regression can be seen in the first column of Table 3.

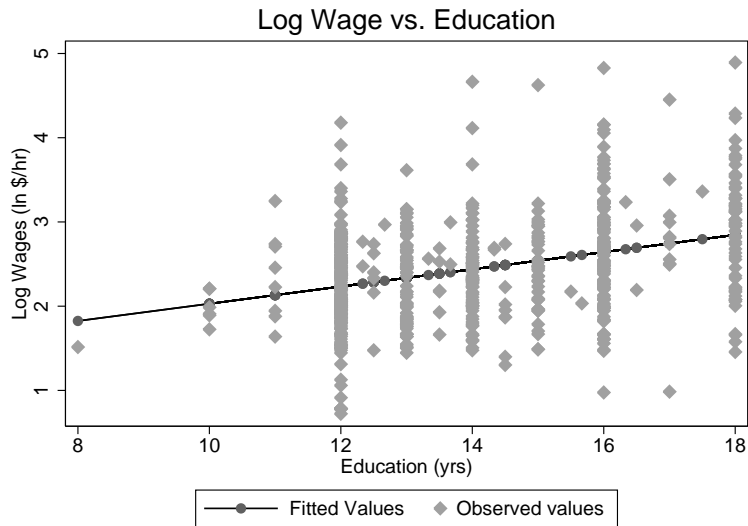


Figure 1: Log Wage vs. Education

The second column of Table 3 contains the result of regressing log wages on education, age, age-squared, and gender and racial dummies. The coefficient on education means that *ceteris paribus* each additional year of education causes an average 11% increase in wages. Likewise, *ceteris paribus* women's wages are on average 32% lower than those of men. Surprisingly, the race coefficient indicates that whites' wages are *ceteris paribus* on average 11% lower than those of blacks, but this coefficient is not significantly different from zero. Since age enters the wage equation quadratically, its interpretation is less straightforward. The estimated equation implies that, *ceteris paribus*, average wages increase with age until the age of 50, and then decline thereafter. This is a plausible life-cycle wage profile, and is

*Unless otherwise noted, all regressions include a constant.

in any case more plausible than a simple linear relationship between age and wages. Evidently including coefficients on age to the third and fourth powers, as in the second column of Table 3, does not significantly improve the fit of the model.

- (b) At 11.0%, the estimated return to education from the multivariate regression is higher than the one from the bivariate regression, which is only 10.2%. This implies that education is not distributed randomly across the twins population or, in other words, that educational attainment may be correlated with other twin characteristics.

Table 1: Descriptive Statistics

| | High School | College | Total |
|----------------|------------------|------------------|------------------|
| Hourly Wage | 11.28 (7.404) | 17.95 (15.59) | 13.82 (11.69) |
| Age | 41.46 (11.48) | 35.39 (10.78) | 39.14 (11.58) |
| Female | 0.629 (0.484) | 0.587 (0.494) | 0.613 (0.488) |
| White | 0.922 (0.267) | 0.877 (0.330) | 0.905 (0.293) |
| Self-Employed | 0.115 (0.310) | 0.130 (0.333) | 0.121 (0.318) |
| Union Member | 0.299 (0.456) | 0.178 (0.381) | 0.253 (0.433) |
| Ever Married | 0.817 (0.388) | 0.475 (0.499) | 0.686 (0.464) |
| Father's Educ. | 10.63 (2.639) | 13.05 (2.879) | 11.57 (2.976) |
| Mother's Educ. | 11.28 (2.125) | 12.84 (2.656) | 11.89 (2.465) |
| <i>N</i> | 224 | 138 | 362 |

Mean values; standard deviations in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The mean characteristics of individuals with high school and college degrees can be seen in Table 1. A statistical comparison of these mean values can be seen in Table 2. As is evident from these tables, individuals with different levels of education exhibit significantly different observable characteristics. There are likely unobserved characteristics that also vary systematically between such groups, such as

cognitive skills or the ability to delay gratification, that are also related to earnings. These considerations imply that the OLS estimate of the relation between log wages and education may be biased.

Table 2: Mean Comparison Tests

| | Difference | |
|----------------|------------|---------|
| Hourly Wage | -6.670*** | (-5.48) |
| Age | 6.073*** | (5.00) |
| Female | 0.0425 | (0.81) |
| White | 0.0451 | (1.42) |
| Self-Employed | -0.0151 | (-0.44) |
| Union Member | 0.122** | (2.62) |
| Ever Married | 0.342*** | (7.30) |
| Father's Educ. | -2.425*** | (-8.12) |
| Mother's Educ. | -1.565*** | (-6.12) |
| N | 362 | |

High School minus College mean values;

t -statistics in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

- (c) Estimated schooling coefficients for the level wage equation can be seen in Figure 2, and corresponding estimates for the log wage equation can be seen in Figure 3.

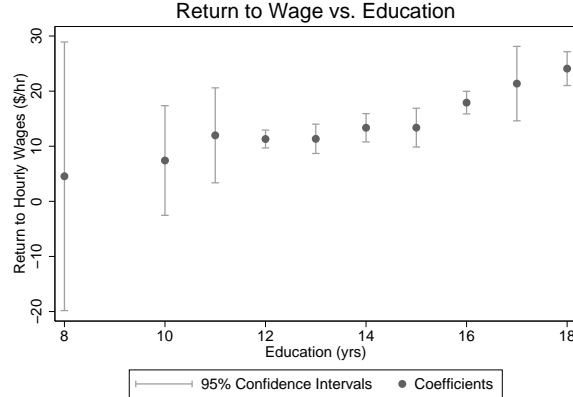


Figure 2: Return to Wage vs. Education

The coefficient estimates from regressing level wages and log wages on dummy variables for each level of schooling can be seen in Figures 2 and 3, respectively. The coefficients in both cases suggest a nonlinear

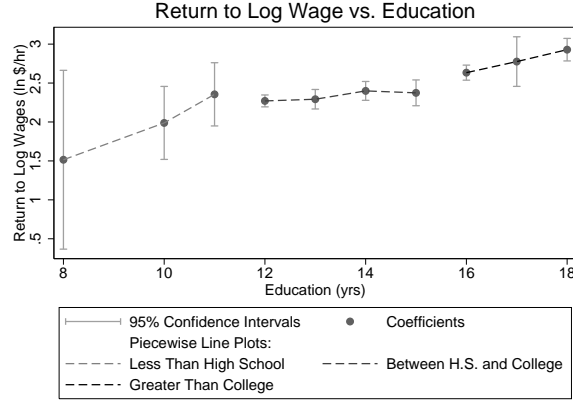


Figure 3: Return to Log Wage vs. Education

relationship between education and wages. This nonlinearity is more pronounced in the log wages equation, which suggests that the log wage profile across education is increasing throughout high school, roughly flat between high school and college, and increasing for the two years thereafter.

The results of a dummy variable regression for log wages that also includes age, age-squared, gender, and race can be seen in the fourth column of Table 3. An analogous regression for level wages can be seen in the sixth column of Table 3. Allowing for nonlinearities in the return to education evidently improves the fit of the model, since the R^2 for each dummy regression is greater than twice that of its parametric counterpart.

- (d) A scatter plot of hourly wages versus education can be seen in Figure 4. Evidently hourly wages exhibit greater variation for higher levels of education, suggesting heteroskedasticity in the level wage regression model. As can be seen in Figure 1, this is not the case for the log wage regression model.
- (e) The result of regressing log wages on education, age, age-squared, and gender and racial dummies can be seen in the second column of Table 4. Eicker-White standard errors are reported in brackets underneath the coefficient estimates. These standard errors are corrected for heteroskedasticity by using the OLS residuals to more accurately estimate the variance of the OLS estimators. In the case of the log wage equations, the Eicker-White standard errors for the education coefficients are numerically equivalent to their OLS counterparts. This provides evidence *against* heteroskedasticity.
- (f) The results of regressing the residuals from both the log and level wage equations on education, age, age-squared, gender, and race can

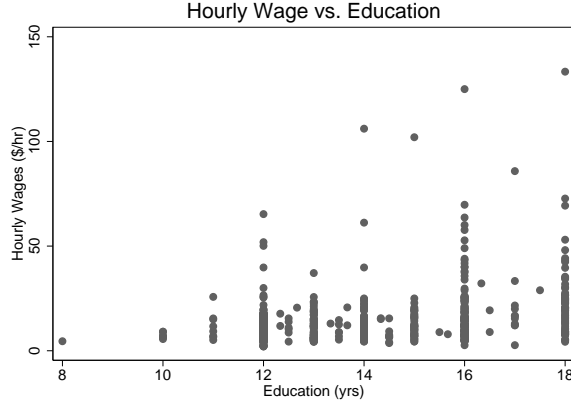


Figure 4: Hourly Wage vs. Education

be seen in columns one and three of Table 5, respectively. In the case of log wages, $nR^2 = 16.32$, and in the case of level wages, $nR^2 = 34.00$. Both are greater than 11.07, which is the 5% critical value for a χ^2 -distribution with 5 degrees of freedom. This provides evidence of heteroskedasticity in both models. With a larger nR^2 , the level wage residuals appear to be more heteroskedastic.

The results of adding education-squared and interaction terms to the aforementioned regressions can be seen in columns two and four of Table 5. For these regressions, in the case of log wages, $nR^2 = 33.32$, and in the case of level wages, $nR^2 = 62.56$. Both are greater than 19.68, which is the 5% critical value for a χ^2 -distribution with 11 degrees of freedom. This once again provides evidence of heteroskedasticity in both models.

- (g) The assumption that the residuals from the log wage regression are uncorrelated may be violated because it is plausible that error terms are correlated within twin pairs, due to common unobservables (such as genetic heritage).

The result of regressing log wages on education, age, age-squared, and gender and racial dummies can be seen in the second column of Table 4. Clustered standard errors are reported in braces underneath the coefficient estimates. The standard errors that account for clustering are larger because they incorporate the “off-diagonal” elements of the residual variance-covariance matrix when estimating the variance of the OLS estimators.

- (h) The result of regressing average log wages on average education, age, age-squared, and gender and racial dummies can be seen in the fourth column of Table 4. This corrects for the clustering in the residuals since each twin pair now has a single error term, which is uncorrelated

with other twin pairs' error terms.

- (i) The bias in the estimated return to education from the regression of log wages on education, age, age-squared, gender, and race is as follows:

$$E[\hat{\beta}_{OLS}] = \left(1 - \frac{\lambda}{1 - R_{s,X}^2}\right) \beta, \quad (1)$$

where $R_{s,X}^2$ is the R^2 from regressing s on the other regressors and λ , the “noise-to-total variance ratio”, is defined as follows:

$$\lambda \equiv \frac{\sigma_u^2}{\sigma_{s^*}^2 + \sigma_u^2} : \frac{\text{Noise}}{\text{Signal} + \text{Noise}} \quad (2)$$

- (j) A Two-Stage Least Squares (2SLS) estimate of the log wage equation, using the other twin's report of an individual's education level as an instrument for self-reported education, can be seen in the second column of Table 6. The 2SLS estimate of the return to education is larger than the OLS estimate presumably because it reduces the bias due to measurement error.

Under the assumption that the measurement errors are classically distributed, the “reliability ratio” is defined as follows:

$$1 - \frac{\sigma_e^2}{\sigma_s^2} = \frac{\text{Var}(s_1^*)}{\sqrt{\text{Var}(s_1^1)\text{Var}(s_1^2)}} = \text{Corr}(s_1^1, s_1^2) \quad (3)$$

The reliability ratio of the education data is roughly 93%.

- (k) If there is an unmeasured factor that is correlated with both an individual's educational attainment and his earnings, then this can bias the OLS estimate of the return to education by introducing a correlation between education and the error term in the regression model. In this case, the “omitted variables bias” is as follows:

$$E[\hat{\beta}_{OLS}] = \beta + \lambda\theta, \quad (4)$$

where λ is the partial correlation coefficient between s and A , and θ is the partial correlation coefficient between log wages and A .

- (l) The result of regressing the difference in wages between twins on their difference in educational attainment can be seen in the third column of Table 6. Note that the coefficient on education is lower in this regression than it is in the result of regressing log wages on education, age, age-squared, and gender and racial dummies. This suggests that omitted variables may be biasing up the latter estimate.
- (m) A scatterplot of the difference in twins' log wages and the difference in twins' self-reported education can be seen in Figure 5. Most of the observations are clustered around 0 on the x -axis, suggesting

that measurement error could be a large component of the variation in this variable. If the latter is the case, then measurement error may be biasing *down* the estimated return to education in the first-differenced equation, contrary to the omitted variables explanation offered in 11.

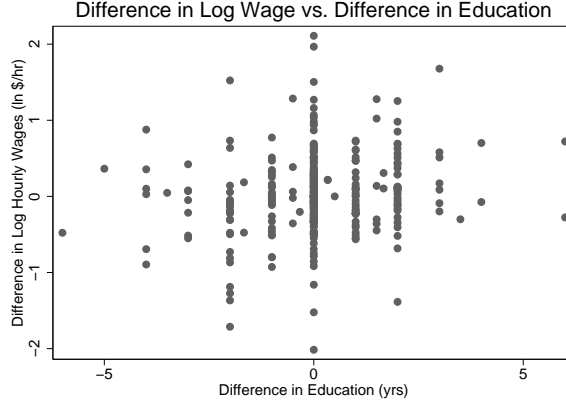


Figure 5: Difference in Log Wage vs. Difference in Education

First-differencing can exacerbate measurement error by subtracting out much of the signal in the differenced variables, so that measurement error becomes a larger component of the remaining variation.

- (n) A 2SLS regression of the the first-differenced log wage equation, using the second twin's reported difference in education as an instrument for the first twin's reported difference in education, can be seen in the fourth column of Table 6. The estimated return to education is now larger than that of part 11, because instrumenting for the difference in education presumably reduces the bias due to measurement error. If the measurement error of a twin's self-report is correlated with the measurement error of his twin's report of his education, then the classical measurement error assumptions are violated and instrumenting for the difference in education as described above no longer eliminates measurement error bias in the estimates. A solution to this problem would be using the difference in cross-reports in education as an instrument for the difference in self-reports.
- (o) If the classical measurement error assumptions hold, one may conclude that the size of the omitted variables bias in the OLS estimate of the returns to education is "small". (Compare the coefficient on education in the first column of Table 6 that of the fourth column.) Comparing twin pair differences across families reduces the omitted variables problem by differencing out unobservables that are common between twins before making cross-family comparisons. However, it

is doubtful that this operation removes all relevant unobservables from the data.

2. (a) The result of regressing log wages on education, experience, experience-squared, and dummies for gender and marital status and computer use, can be seen in the first column of Table 7. The coefficient on education means that *ceteris paribus* each additional year of education causes an average 7% increase in wages. Likewise, *ceteris paribus* women's wages are on average 22% lower than those of men, and *ceteris paribus* married individuals make on average 3.5% more than those that are unmarried. The coefficients on experience suggest that *ceteris paribus* the average returns to experience are increasing at a decreasing rate. Finally, *ceteris paribus* the coefficient on computer use implies an average 17% wage premium for using a computer.
- (b) The result of adding dummies for pencil, telephone, calculator, and hammer use to the regression described in 2a can be seen in the second column of Table 7. This leads to slight reductions in the coefficients on education and computer use. The result of adding dummies for individuals' occupation categories to this regression can be seen in the third column of Table 7. This leads to a significant reduction in the coefficient on computer use, reducing it by more than half of its value in 2a. This suggests that there is substantial omitted variable bias in the initial estimated return to computer use, since controlling for additional factors reduces that estimate by so much. In other words, the initial estimate is "soaking up" the effects of occupation-specific factors that are positively correlated with wages (and computer use).

Table 3: OLS Estimates of Wage Equations

| | Log Wages | | | | Level Wages | |
|------------------|--------------------------------|---------------------------------|---------------------------------|---------------------------------|----------------------------------|---------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Constant | 1.008*** (0.153) [0.158] | -1.095*** (0.261) [0.292] | -3.937* (1.832) [2.097] | | -36.812*** (5.937) [6.481] | |
| Education | 0.102*** (0.011) [0.011] | 0.110*** (0.010) [0.010] | 0.109*** (0.010) [0.010] | | 2.077*** (0.217) [0.281] | |
| Age | | 0.104*** (0.010) [0.012] | 0.414* (0.194) [0.219] | 0.105*** (0.011) [0.013] | 1.077*** (0.239) [0.298] | 1.057*** (0.242) [0.317] |
| Age ² | | -0.001*** (0.000) [0.000] | -0.013 (0.007) [0.008] | -0.001*** (0.000) [0.000] | -0.010*** (0.003) [0.004] | -0.009** (0.003) [0.004] |
| Age ³ | | | 0.000 (0.000) [0.000] | | | |
| Age ⁴ | | | -0.000 (0.000) [0.000] | | | |
| Female | | -0.318*** (0.040) [0.040] | -0.317*** (0.040) [0.041] | -0.322*** (0.040) [0.040] | -5.896*** (0.910) [0.948] | -6.002*** (0.920) [0.970] |
| White | | -0.100 (0.072) [0.068] | -0.110 (0.073) [0.069] | -0.098 (0.073) [0.070] | -0.283 (1.641) [1.376] | -0.269 (1.658) [1.404] |
| Educ. Dummies | No | No | No | Yes | No | Yes |
| R^2 | 0.116 | 0.339 | 0.341 | 0.960 | 0.223 | 0.653 |
| F | 89.322 | 69.058 | 49.764 | 1141.595 | 38.691 | 89.333 |
| N | 680 | 680 | 680 | 680 | 680 | 680 |

Standard errors in parentheses; robust standard errors in brackets.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Robust OLS Estimates of Wage Equations

| | Log Wages | | Log Wages | |
|------------------|---|--|--------------------------------|---------------------------------|
| | (1) | (2) | (3) | (4) |
| Constant | 1.008*** (0.153) [0.158] {0.191} | -1.095*** (0.261) [0.292] {0.339} | 0.926*** (0.209) [0.211] | -1.229*** (0.330) [0.359] |
| Education | 0.102*** (0.011) [0.011] {0.014} | 0.110*** (0.010) [0.010] {0.013} | 0.108*** (0.015) [0.015] | 0.118*** (0.013) [0.014] |
| Age | | 0.104*** (0.010) [0.012] {0.015} | | 0.105*** (0.013) [0.015] |
| Age ² | | -0.001*** (0.000) [0.000] {0.000} | | -0.001*** (0.000) [0.000] |
| Female | | -0.318*** (0.040) [0.040] {0.049} | | -0.317*** (0.049) [0.050] |
| White | | -0.100 (0.072) [0.068] {0.069} | | -0.092 (0.091) [0.072] |
| R^2 | 0.116 | 0.339 | 0.137 | 0.408 |
| F | 89.322 | 69.058 | 53.806 | 46.049 |
| N | 680 | 680 | 340 | 340 |

Standard errors in parentheses; robust standard errors in brackets;
clustered standard errors in braces.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: OLS Estimates of Residual Wage Equations

| | Log Wages | | Level Wages | |
|-------------------|--------------------|---------------------|-------------------------|-------------------------|
| | (1) | (2) | (3) | (4) |
| Constant | 0.198 (0.252) | 1.321 (1.374) | -484.849 (363.967) | 2209.863 (1960.934) |
| Education | 0.021* (0.009) | -0.153 (0.151) | 47.157*** (13.316) | -156.828 (215.834) |
| Educ ² | | 0.005 (0.005) | | 1.236 (6.754) |
| Age | -0.014 (0.010) | -0.026 (0.018) | -8.443 (14.627) | -71.031** (24.990) |
| Age ² | 0.000 (0.000) | 0.000 (0.000) | 0.192 (0.176) | 0.189 (0.178) |
| Female | -0.095* (0.039) | 0.581 (0.317) | -184.726*** (55.773) | 1447.551** (452.086) |
| White | 0.048 (0.070) | 0.401 (0.661) | 89.816 (100.606) | -856.931 (943.218) |
| Educ · Age | | 0.001 (0.001) | | 4.460*** (1.169) |
| Female · Age | | 0.002 (0.003) | | -9.934* (4.852) |
| Female · Educ | | -0.053** (0.019) | | -87.799** (27.200) |
| White · Age | | -0.011 (0.006) | | 5.925 (8.718) |
| White · Educ | | 0.003 (0.037) | | 45.280 (52.939) |
| R^2 | 0.024 | 0.049 | 0.050 | 0.092 |
| F | 3.362 | 3.159 | 7.032 | 6.139 |
| N | 680 | 680 | 680 | 680 |

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: OLS, 2SLS, and 1st-Diff. Estimates of Log Wage Equations

| | (1) OLS | (2) 2SLS | (3) 1 st -Diff. | (4) 1 st -Diff. 2SLS |
|------------------|--|--|-------------------------------|------------------------------------|
| Constant | −1.095*** (0.261) [0.292] {0.339} | −1.188*** (0.269) [0.306] {0.354} | | |
| Education | 0.110*** (0.010) [0.010] {0.013} | 0.116*** (0.010) [0.011] {0.014} | 0.062** (0.019) [0.020] | 0.108*** (0.030) [0.034] |
| Age | 0.104*** (0.010) [0.012] {0.015} | 0.104*** (0.011) [0.012] {0.015} | | |
| Age ² | −0.001*** (0.000) [0.000] {0.000} | −0.001*** (0.000) [0.000] {0.000} | | |
| Female | −0.318*** (0.040) [0.040] {0.049} | −0.316*** (0.040) [0.040] {0.049} | | |
| White | −0.100 (0.072) [0.068] {0.069} | −0.098 (0.072) [0.068] {0.068} | | |
| R^2 | 0.339 | 0.338 | 0.031 | . |
| F | 69.058 | 67.581 | 10.880 | . |
| N | 680 | 680 | 340 | 340 |

Own-reports of education instrumented for with twin-reports.

Standard errors in parentheses; robust standard errors in brackets;
clustered standard errors in braces.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: OLS Estimates of Wage Equations

| | (1) | (2) | (3) |
|-------------------------|---------------------------------|---------------------------------|---------------------------------|
| Constant | 1.701*** (0.019) [0.020] | 1.751*** (0.020) [0.021] | |
| Education | 0.070*** (0.001) [0.001] | 0.065*** (0.001) [0.001] | 0.040*** (0.002) [0.002] |
| Experience | 0.030*** (0.001) [0.001] | 0.029*** (0.001) [0.001] | 0.026*** (0.001) [0.001] |
| Experience ² | -0.046*** (0.002) [0.002] | -0.044*** (0.002) [0.002] | -0.038*** (0.002) [0.002] |
| Female | -0.215*** (0.006) [0.006] | -0.229*** (0.006) [0.006] | -0.160*** (0.008) [0.008] |
| Married | 0.035*** (0.006) [0.006] | 0.033*** (0.006) [0.006] | 0.034*** (0.006) [0.006] |
| Computer | 0.172*** (0.006) [0.006] | 0.120*** (0.007) [0.007] | 0.068*** (0.007) [0.007] |
| Pencil | | 0.031*** (0.008) [0.008] | 0.007 (0.008) [0.008] |
| Telephone | | 0.042*** (0.008) [0.008] | 0.048*** (0.008) [0.008] |
| Calculator | | 0.046*** (0.007) [0.007] | 0.022** (0.007) [0.007] |
| Hammer | | -0.035*** (0.006) [0.006] | -0.021** (0.008) [0.008] |
| Occ. Dummies | No | No | Yes |
| R^2 | 0.333 | 0.343 | 0.987 |
| F | 1669.843 | 1047.286 | 1295.686 |
| N | 20042 | 20042 | 20042 |

Standard errors in parentheses; robust standard errors in brackets.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$