Econometrics

Exercise: Omitted Variable Bias

Lets take a closer look at the OVB.

Specifically say that wages are determined by: $log(Wage) = \beta_0 + \beta_1 Educ + \beta_2 IQ + u$
(So that $E(u|Educ, IQ) = 0$)

But suppose we don't have data on IQ and so instead run the regression:
$log(Wage) = \alpha_0 + \alpha_1 Educ + v$

Note now that: $v = \beta_2 IQ + u$

We saw the OVB formula of: $\hat{\alpha}_1 \to \beta_1 + \rho_{x_1 v} \frac{\sigma_v}{\sigma_{X_1}}$

Let δ_1 be the coefficient on $Educ$ from a regression of IQ on $Educ$: $IQ = \delta_0 + \delta_1 Educ + \epsilon$

TURN IN

Show the OVB formula implies: $\hat{\alpha}_1 \to \beta_1 + \beta_2 \delta_1$

Hint:
1. Rewrite the correlation in the second part of the OVB formula in terms of covariance/variances
2. Cancel terms
3. Rewrite the $cov(X_1, v)$ in terms of IQ and u
(recall that $cov(Y, aX + bZ) = a \cdot cov(Y, X) + b \cdot cov(Y, Z)$)
4. Split up the terms
5. Note one term is the definition of the simple regression of IQ on $Educ$
6. Note the other term is zero because $E(u|Educ) = 0$ implying zero correlation

Now show this works with data:

1. Load the data set 'wages.csv'.
2. Regress IQ on $Educ$ to get $\hat{\delta_1}$ from $IQ = \delta_0 + \delta_1 Educ + \epsilon$
3. Regress log wages on to education to get $\hat{\alpha_1}$ from $log(Wage) = \alpha_0 + \alpha_1 Educ + v$
4. Run multiple regression: $log(Wage) = \beta_0 + \beta_1 Educ + \beta_2 IQ + u$
5. Verify that $\hat{\alpha}_1 = \hat{\beta_1} + \hat{\beta_2}\hat{\delta_1}$

Copy your code and outputs from R and save in a word file.
