Quantitative Economics Problem Set 3

1a The estimated coefficient on sibs is -0.094. On average, ~~a man with one more sibling has~~ having one more sibling is associated with 0.094 fewer years of education. holding meduc and feduc constant

That the estimated coefficient is negative is not entirely surprising. Plausibly, ~~there~~ the families of men with a greater number of siblings shared a finite amount of financial and educational resources between a larger number of children, hence such men received less investment from their parents in their education, and so had fewer total years in education.

On average, having $1/0.094 = 10.632 \approx 11$ more siblings is associated with having 1 ~~fewer~~ less year of education.

b On average, a man whose mother had one more year in education, holding constant the number of his siblings and his father's total years in education, had 0.131 more total years of education.

Other causal determinants of educ that are unobserved and collected in the population regression residual include country of residence, area of residence (urban / rural), household income, and household wealth. These ~~determina~~ determinants are not plausibly uncorrelated with meduc. ~~A~~ The mother of a man who lived in a developed country and in an urban area ~~with greater access to~~ is more likely to have grown up in such a country and area, hence more likely to have had greater access to education and more likely to have higher total years in education. ~~A mother~~ A mother with high total years in education is also more likely to have higher income, a man belonging to such a household would have had higher household income.

Then, the unobserved determinants of educ are ~~not~~ correlated with the observed determinants. Orthogonality fails in the causal model, the population regression parameters do not coincide with the causal model parameters. The sample regression estimates are consistent for the former but not the latter, hence do not provide a reliable estimate for how a mother's educational attainment influences that of her son.

Further, the estimated coefficient provides a reliable estimate of the causal effect of interest only if that causal effect is linear. We have no reason to think so in this case.

c $\hat{E}(educ_A \mid sibs_A = 0, meduc_A = feduc_A = 12)$
$= 10.36 - 0.094(0) + 0.131(12) + 0.210(12) + \hat{E}\hat{u}_A$
$= 14.452$

$\hat{E}(educ_B \mid sibs_B = 0, meduc_B = feduc_B = 16)$
$= 10.36 - 0.094(0) + 0.131(16) + 0.210(16) + \hat{E}\hat{u}_B$
$= 15.816$

Predicted difference $= 15.816 - 14.452 = 1.364$

An intergenerational ~~regression to the mean~~ mean reversion is observed because the ~~effect of an increase in meduc~~ ~~and~~ higher meduc and higher feduc are, on average, associated with ~~a less than one to one~~ ~~one to one~~ higher educ, but in a less than one-to-one proportion.

d Other causal determinants and their correlation with the observed determinants are discussed in (b).

2a Age is a causal determinant of wage. Older workers tend to have greater knowledge and experience, hence greater productivity, and higher wage. Then, including age in the regression model reduces the magnitude hence the variance of the residuals, hence increases the precision of the estimates of the other regression coefficients.

On average, a worker who is one year older has an hourly wage that is 0.51 dollars higher, holding degree status and gender constant.

The coefficient on age consistently estimates the causal effect of age on hourly wage only if age is uncorrelated with other unobserved determinants of wage. This is not plausible. The economic conditions at the time that a person entered the workforce is ~~a~~ one such unobserved determinant of wage, and is highly correlated with ~~the~~ age. Then, the ~~the~~ population regression coefficient on ~~the~~ age does not coincide with ~~that of the~~ the causal model coefficient, and the sample regression coefficient is consistent for the former but not the latter.

KOKUYO LOOSE-LEAF 7-6816S-5

b On average, female workers had an hourly wage 3.81 less than that of male workers, holding degree status and age constant.

The negative coefficient constitutes decisive evidence that being female has a negative causal effect (due to discrimination) on wages only if gender is uncorrelated with other unobserved ~~deter~~ determinants of wage. This is not entirely plausible. ~~If~~ One such determinant is sector of employment and ~~a worker's role job respon~~ the nature of work. If it is the case that female workers disproportionately seek employment in sectors or lines of work that pay less, then the negative coefficient on female ~~coeff~~ is not a consistent estimator of the ~~effect~~ causal effect of gender on wage, and does not constitute decisive evidence ~~of~~ of discrimination.

c No. degree is not plausibly uncorrelated with the other unobserved determinants of wage. One such determinant is general cognitive ability, another is social capital. Workers with high general cognitive ability are more likely to have a degree because admission into and graduation from a degree programme requires some level of cognitive ability. Workers with high social capital are more likely to have a degree because attending a degree programme is costly and ~~the~~ the ability to afford such a programme is likely correlated with parents' social capital. Then, the population regression (of wage on degree, female, and age) parameters do not coincide with the causal model parameters, and the sample regression coefficients are consistent for the former but not the latter.

Supposing that a worker is not ~~directly~~ employed in some sort of family business, parental income is not plausibly a causal determinant of wage. ~~Then, parental income~~ Parental income is presumably correlated with a young worker's social capital. Then, parental income can be used as a proxy for social capital.

There is no apparent proxy for general cognitive ability apart from test scores, but only supposing that these are not known by employers hence not a causal determinant of wage. It is also necessary to suppose that the above proxies are uncorrelated with degree status. Admittedly, this is not very plausible. Supposing that such plausible proxies exist, regression of wage on ~~female~~ degree, female, age, and these proxies yields a consistent estimate of the causal effect of degree on wage.

3a No, the regressors are perfectly multicollinear. ~~As~~ An increase (decrease) in study is necessarily associated with a one-to-one decrease (increase) in the sum of sleep, work, and leisure. Then it does not make sense to speak of a causal effect of study because ~~so that~~ there is no variation in study as such but only reallocations of time from sleep, work, or leisure to study. It does not make sense to hold sleep, work, and leisure constant while varying study.

b The given causal model cannot be estimated by OLS regression. The residual in each auxiliary regression is zero because of perfect multicollinearity. For example, in the auxiliary regression $study = \pi_0 + \pi_1 sleep + \pi_2 work + \pi_3 leisure + \tilde{study}$, OLS regression yields the parameters $\hat{\pi}_0 = 168$, $\hat{\pi}_1 = \hat{\pi}_2 = \hat{\pi}_3 = -1$, $\hat{\tilde{study}} = 0$. Then, the sample variance of the residual is zero. Hence the OLS regression coefficient which is equal to the sample covariance of score and the residual divided by the variance of the residual is undefined.

c Omit one of the perfectly multicollinear regressors. Then, the coefficients ~~of~~ on the remaining regressors can be interpreted as the average difference in score associated with a higher ~~one~~ value of that regressor and a lower (in one-to-one proportion) value of the omitted regressor. For example, if study is omitted, the coefficient on sleep is the average difference in score associated with having ~~higher sleep~~ one hour more sleep and one hour less study.

Having omitted one regressor, the ~~coeff~~ residual in ~~the~~ each auxiliary regression is non-zero and has non-zero sample variance (supposing there is no perfect multicollinearity between the remaining regressors in the data, which could ~~be the~~ fail to hold if, for example, work = 0 for all observations in the data). Then, the parameters of the ~~population~~ causal model can be estimated (not necessarily consistently) by OLS regression.

4 Population linear regression problem:

$$\min_{b_0,b_1,b_2} E(Y-b_0-b_1X-b_2X_2)^2$$
$$= \min_{b_0,b_1,b_2} E(\beta_0+\beta_1X_1+\beta_2X_2+u-b_0-b_1X_1-b_2X_2)^2$$
$$= \min_{b_0,b_1,b_2} E([\beta_0-b_0]+[\beta_1-b_1]X_1+[\beta_2-b_2]X_2+u)^2$$

Let $\delta_0=\beta_0-b_0$, $\delta_1=\beta_1-b_1$, $\delta_2=\beta_2-b_2$ for notational convenience

$$= \min_{b_0,b_1,b_2} E(\delta_0+\delta_1X_1+\delta_2X_2+u)^2$$
$$= \min_{b_0,b_1,b_2} E(\delta_0^2+\delta_0\delta_1X_1+\delta_0\delta_2X_2+\delta_0u + \delta_1X_1\delta_0+\delta_1^2X_1^2+\delta_1X_1\delta_2X_2+\delta_1X_1u + \delta_2X_2\delta_0+\delta_2X_2\delta_1X_1+\delta_2^2X_2^2+\delta_2X_2u)$$
$$= \min_{b_0,b_1,b_2} [\delta_0^2 + (\delta_0\delta_1+\delta_1\delta_0)EX_1 + (\delta_0\delta_2+\delta_2\delta_0)EX_2 + \delta_1^2EX_1^2 + \delta_2^2EX_2^2 + (\delta_1\delta_2+\delta_2\delta_1)EX_1X_2 + 2\delta_0Eu + 2\delta_1EX_1u + 2\delta_2EX_2u + Eu^2$$
$$= \min_{b_0,b_1,b_2} [\delta_0^2 + 2\delta_0\delta_1EX_1 + 2\delta_0\delta_2EX_2 + \delta_1^2EX_1^2 + \delta_2^2EX_2^2 + 2\delta_1\delta_2EX_1X_2 + Eu^2]$$

FOC δ_0: $2\delta_0+2\delta_1EX_1+2\delta_2EX_2=0$
FOC δ_1: $2\delta_0EX_1+2\delta_1EX_1^2+2\delta_2EX_1X_2=0$
FOC δ_2: $2\delta_0EX_2+2\delta_2EX_2^2+2\delta_1EX_1X_2=0$

By inspection, $\delta_0=\delta_1=\delta_2=0$ satisfies the FOC.

SOC δ_0: $2>0$
SOC δ_1: $2EX_1^2>0$
SOC δ_2: $2EX_2^2>0$

$\delta_0=\delta_1=\delta_2=0$ satisfies the SOC for a minimum, $b_0,b_1,b_2=\beta_0,\beta_1,\beta_2$ solves the population linear regression problem.

5a Given $E[u|x_1,x_2]=0$

$$Eu = E(E[u|x_1,x_2]) = E(0) = 0$$

$$\begin{aligned}\text{Cov}(x_1,u) &= E(x_1-Ex_1)(u-Eu)\\ &= E(x_1-Ex_1)(u)\\ &= E(ux_1-uEx_1)\\ &= E(ux_1)-EuEx_1\\ &= E[E(ux_1|x_1,x_2)]\\ &= E[x_1E(u|x_1,x_2)]\\ &= E[0]\\ &= 0\end{aligned}$$

By an analogous argument, $\text{Cov}(x_2,u)=0$

Then, the given equation coincides with a population linear regression of Y on x_1,x_2.

b β_1 is equal to the coefficient on x_1 in a population linear regression of Y on x_1 alone only if orthogonality holds in the corresponding causal model, i.e. only if $E(x_1(\beta_2x_2+u))=0$.

$E(x_1(\beta_2x_2+u)) = \beta_2Ex_1x_2 = 0 \iff Ex_1x_2=0$. This condition is not necessarily satisfied.

6 $$\begin{aligned}E[\textstyle\sum_{i=1}^n x_iu_i \mid x_1,\dots,x_n] &= \textstyle\sum_{i=1}^n E[x_iu_i|x_1,\dots,x_n]\\ &= \textstyle\sum_{i=1}^n x_iE[u_i|x_1,\dots,x_n]\\ &= \textstyle\sum_{i=1}^n x_iE[u_i|x_i]\\ &= \textstyle\sum_{i=1}^n x_i(0)\\ &= 0\end{aligned}$$

where the first equality follows by linearity of conditional expectation, the second follows by conditioning, the third by independence, the fourth by substitution, and the fifth by basic operations.

b $$\begin{aligned}&E[\textstyle\sum_{i=1}^n (x_i-\bar{x})(u_i-\bar{u}) \mid x_1,\dots,x_n]\\ &= \textstyle\sum_{i=1}^n E[(x_i-\bar{x})(u_i-\bar{u})|x_1,\dots,x_n]\\ &= \textstyle\sum_{i=1}^n E[x_iu_i-\bar{u}x_i-\bar{x}u_i+\bar{x}\bar{u}|x_1,\dots,x_n]\\ &= \textstyle\sum_{i=1}^n E[x_iu_i|x_1,\dots,x_n] - \bar{u}x_i - \bar{x}E[u_i|x_1,\dots,x_n] + \bar{x}\bar{u}\\ &= \textstyle\sum_{i=1}^n 0-\bar{u}x_i-0+\bar{x}\bar{u}\\ &= \bar{u}\textstyle\sum_{i=1}^n x_i - \bar{x}\\ &= \bar{u}\textstyle\sum_{i=1}^n (x_i - \frac{1}{n}\sum_{j=1}^n x_j)\\ &= \bar{u}\textstyle\sum_{i=1}^n x_i - \bar{u}n\frac{1}{n}\sum_{j=1}^n x_j\\ &= 0\end{aligned}$$

where the first equality follows by linearity, the second by expansion, the third by linearity, and conditioning, the fourth by substitution and independence, and the remaining by basic operations.

c $$\begin{aligned}\hat{\beta}_1 &= \widehat{\text{cov}}(x,Y)/\widehat{\text{var}}(x)\\ &= \widehat{\text{cov}}(x,\beta_0+\beta_1x+u)/\widehat{\text{var}}(x)\\ &= [\beta_1\widehat{\text{var}}(x)+\widehat{\text{cov}}(x,u)]/\widehat{\text{var}}(x)\\ &= \beta_1+\widehat{\text{cov}}(x,u)/\widehat{\text{var}}(x)\end{aligned}$$

where the first equality follows by construction, the second by substitution, the third by linearity of covariance, the fourth by basic arithmetic.

d From (b)

$$\begin{aligned}&E[\textstyle\sum_{i=1}^n (x_i-\bar{x})(u_i-\bar{u}) \mid x_1,\dots,x_n] = 0\\ \Rightarrow\ &E[\tfrac{1}{n}\textstyle\sum_{i=1}^n (x_i-\bar{x})(u_i-\bar{u})] = 0\\ \Rightarrow\ &E[\widehat{\text{cov}}(x,u)] = 0\\ \Rightarrow\ &E\hat{\beta}_1 = E\beta_1 + E[\widehat{\text{cov}}(x,u)/\widehat{\text{var}}(x)]\\ &= E\beta_1\\ &= \beta_1\end{aligned}$$

KOKUYO LOOSE-LEAF A-S816S-5

7a $\hat{\beta}_1$ construction, $\hat{\beta}_1 = \hat{cov}(Y,X)/\hat{var}(X)$ and $\hat{\beta}_1^* = \hat{cov}(Y^*,X^*)/\hat{var}(X^*)$

$\hat{\beta}_1^* = \hat{cov}(aY+c, bX)/\hat{var}(bX)$
$= ab\hat{cov}(Y,X)/b^2\hat{var}(X)$
$= a/b\ \hat{cov}(Y,X)/\hat{var}(X)$
$= a/b\ \hat{\beta}_1$

$\hat{\beta}_0^* = \hat{E}Y^* - \hat{\beta}_1^*\hat{E}X^*$
$= \hat{E}(aY+c) - a/b\,\hat{\beta}_1\hat{E}(bX)$
$= c + a\hat{E}Y - a\hat{\beta}_1\hat{E}X$
$= c + a\hat{\beta}_0$

b The SER is the square root of the sample variance of the ~~residuals~~ OLS residual.

$\hat{u} = Y - \hat{\beta}_0 - \hat{\beta}_1 X$
$\hat{u}^* = Y^* - \hat{\beta}_0^* - \hat{\beta}_1^* X^*$
$= aY + c - (c + a\hat{\beta}_0) - a/b\,\hat{\beta}_1 bX_1$
$= aY - a\hat{\beta}_0 - a\hat{\beta}_1 X_1$
$= a\hat{u}$

$s^2_{\hat{u}} = n/(n-k-1)\ 1/n\sum_{i=1}^n (\hat{u}_i - 1/n\sum_{i=1}^n \hat{u}_i)^2$
$= 1/(n-k-1)\ n/(n-k-1)\ \hat{E}(\hat{u} - E\hat{u})^2$

$s^2_{\hat{u}^*} = n/(n-k-1)\ \hat{E}(\hat{u}^* - E\hat{u}^*)^2$
$= n/(n-k-1)\ \hat{E}(a\hat{u} - Ea\hat{u})^2$
$= n/(n-k-1)\ a^2\hat{E}(\hat{u} - E\hat{u})^2$
$= a^2 s^2_{\hat{u}}$

$s_{\hat{u}^*} = a s_{\hat{u}}$

The SER is an absolute measure of the variability of Y around fitted values $\hat{Y}$ and so scales in direct proportion to the scaling of Y.

c $SSR = \sum_{i=1}^n \hat{u}_i^2$
$SSR^* = \sum_{i=1}^n \hat{u}^{*2}$
$= \sum_{i=1}^n (a\hat{u})^2$
$= a^2 SSR$

$TSS = \sum_{i=1}^n (Y_i - \hat{E}Y_i)^2$
$TSS^* = \sum_{i=1}^n (Y_i^* - \hat{E}Y_i^*)^2$
$= \sum_{i=1}^n (aY_i + c - \hat{E}(aY_i + c))^2$
$= \sum_{i=1}^n (aY_i + c - a\hat{E}Y_i - c)^2$
$= a^2\sum_{i=1}^n (Y_i - \hat{E}Y_i)^2$
$= a^2 TSS$

$R^2 = 1 - SSR/TSS$
$R^{2*} = 1 - SSR^*/TSS^*$
$= 1 - SSR/TSS$
$= R^2$

R^2 is a relative measure of the variability of Y around fitted values $\hat{Y}$, hence R^2 is invariant to scaling (and/or translation) of Y and/or X.

8 Let the following denote the OLS regression model of X_1 on X_2

$X_1 = \hat{\pi}_0 + \hat{\pi}_2 X_2 + \hat{\tilde{X}}_1$
where $\hat{E}(\hat{\tilde{X}}_1) = \hat{E}(X_2\hat{\tilde{X}}_1) = 0$

Then, by construction, $\hat{\pi}_0 = \hat{E}X_1 - \hat{\pi}_2\hat{E}X_2$ and $\pi_2 = \hat{cov}(X_1,X_2)/\hat{var}(X_2)$

By substitution into the OLS regression model of Y on X_1 and X_2

$Y = \hat{\beta}_0 + \hat{\beta}_1(\hat{\pi}_0 + \hat{\pi}_2 X_2 + \hat{\tilde{X}}_1) + \hat{\beta}_2 X_2 + \hat{u}$
$= (\hat{\beta}_0 + \hat{\beta}_1\hat{\pi}_0) + \hat{\beta}_1\hat{\tilde{X}}_1 + (\hat{\beta}_2 + \hat{\beta}_1\hat{\pi}_2)X_2 + \hat{u}$
where $\hat{E}\hat{\tilde{X}}_1\hat{u} = 0$ given that $\hat{\tilde{X}}_1$ is a linear function of X_1 and X_2 and each of X_1 and X_2 is orthogonal to u, and $\hat{E}X_2\hat{u} = \hat{E}\hat{\tilde{X}}_1\hat{u} = 0$ by construction of the OLS regression model of Y on X_1, X_2. Then,

$Y = \hat{\delta}_0 + \hat{\beta}_1 X_1 + \hat{\delta}_2 X_2 + \hat{u}$
where $\hat{\delta}_0 = \hat{\beta}_0 + \hat{\beta}_1\hat{\pi}_0$ and $\hat{\delta}_2 = \hat{\beta}_2 + \hat{\beta}_1\hat{\pi}_2$ is a OLS regression model of Y on $\hat{\tilde{X}}_1$ and X_2,

Hence, $\hat{\beta}_1$ satisfies $\hat{\beta}_1 = \hat{cov}(Y, \hat{\tilde{X}}_1)/\hat{var}(\hat{\tilde{X}}_1)$.