Quantitative Economics Paper 170524

1a $y_i = \alpha + \varepsilon_i$
$y_i = \beta x_i + \eta_i$

Given that the above equations are population linear regressions, by construction $E\varepsilon_i = 0$ and $E\eta_i = \mathrm{cov}(x_i, \eta_i) = 0$.

$\hat{\alpha}$ solves the sample ~~pr~~ linear regression problem
$\min_\alpha \hat{E}(y_i - \alpha)^2$
FOC: $\partial/\partial\alpha\ \hat{E}(y_i-\alpha)^2$
$= \hat{E}\ 2(y_i-\alpha)(-1)$
$= -2\hat{E}\ (y_i-\alpha)$
$= 0$
$\Rightarrow \hat{E}(y_i-\alpha) = 0$
$\Rightarrow \alpha = \hat{E}y_i = 1/n \sum_{i=1}^n y_i$

Alternatively, $\hat{\alpha}$ is such that ~~E~~ $\hat{E}\hat{\varepsilon}_i = 0$. Then $\hat{E}y_i = \hat{E}(\hat{\alpha} + \hat{\varepsilon}_i) = \hat{\alpha} + \hat{E}\hat{\varepsilon}_i = \hat{\alpha}$.

b $\hat{\beta}$ solves the sample linear regression problem
$\min_\beta \hat{E}(y_i - \beta x_i)^2$
FOC: $\partial/\partial\beta\ \hat{E}(y_i - \beta x_i)^2$
$= \hat{E}\ 2(y_i - \beta x_i)(-x_i)$
$= -2\hat{E}(x_i y_i - \beta x_i^2)$
$= 0$
$\Rightarrow \hat{E}(x_i y_i) - \beta\hat{E}(x_i^2) = 0$
$\Rightarrow \beta = \hat{E}x_i y_i / \hat{E}x_i^2 = 1/n \sum_{i=1}^n x_i y_i \,/\, 1/n \sum_{i=1}^n x_i^2$

2a x_t and y_t are cointegrated with cointegrating coefficient θ iff equilibrium error $\hat{\varepsilon}_t = x_t - \theta y_t$ is stationary.

It is not necessarily the case that if x_t and y_t are non-stationary and have a common stochastic trend that there must exist a cointegrating relationship between x_t and y_t. Consider the following counterexample where each of x_t and y_t also has a deterministic trend, and this is not common to the two.

$x_t = \alpha t + \sum_{s=0}^t \varepsilon_s + u_t$
$y_t = \beta t + \gamma \sum_{s=0}^t \varepsilon_s \quad v_t$

where ε_s, u_t, v_t are independent and iid. Suppose also that $\beta/\alpha \neq \gamma$.

For $\theta = 1/\gamma$, $\varepsilon_t = (\alpha - \beta/\gamma)t + u_t - \theta v_t$. This has a deterministic trend, so it is non-stationary, so x_t and y_t are not cointegrated with $\theta = 1/\gamma$

For $\theta \neq 1/\gamma$, $\varepsilon_t = (\alpha - \beta\theta)t + (1-\gamma\theta)\sum_{s=0}^t \varepsilon_s + u_t - \theta v_t$. This has a stochastic trend, so it is non-stationary, so x_t and y_t are not cointegrated with $\theta \neq 1/\gamma$.

So x_t and y_t are not cointegrated.

b

3a $R^2 = ESS/TSS = 1 - SSR/TSS = 1 - 13.42/73.01 = 0.81619$

The regression has a causal interpretation iff ~~the regr~~ each regressor is exogenous, this is iff there is no omitted variable bias, simultaneity, and/or measurement error. Whether such sources of endogeneity are ~~present~~ present is largely independent of the R^2 of a regression. R^2 is a measure of fit, i.e. the fraction of the variance in the dependent variable that can be descriptively or statistically accounted for by variance in the regressors.

A regression with a causal ~~determin~~ interpretation could have very poor fit and very low R^2 if the causal relationship is ~~we~~ weak. For example, a regression of exam results on level of caffeine in a student's body during the exam will likely have a low R^2 because caffeine levels have a very weak causal relationship to exam scores.

In contrast, a regression with no causal interpretation could have a high R^2. For example, regression of scores ~~of~~ on this exam on scores on the microeconomics exam will likely have a high R^2, but no causal interpretation. ~~because~~

b On average, a parent's having higher log income by 1 is associated with a child's having higher log income by 0.568, holding constant whether the child is living in an urban or rural area. Equivalently, on average, a parent's having $e-1 = 1.7183$ times higher income is associated with a child's having $e^{0.568}-1 = 0.76473$ times higher income, so twofold higher parental income is ~~as~~ ~~is~~ associated with $e^{0.568 \times \ln 2} = 1.4825$ -fold higher child's income, holding urban/rural constant.

$H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

t-statistic
$t = (\hat{\beta} - \beta)/se(\hat{\beta}) = (0.568 - 0)/0.005 = 113.6$

Under the null, given a sufficiently large iid random sample, by CLT, $t \overset{a}{\sim} N(0,1)$.

Reject the null if $|t| > c_\alpha$, where c_α is the critical value drawn from the standard normal distribution at the level of significance α.

~~Under, T~~

The p-value is the probability under the null of observing a t-statistic as unfavourable to the null as that actually observed.

$p = P(|t| > t^{act})$
$= P(|N(0,1)| > 113.6)$
$= 2\Phi(-113.6)$
$= 0.000$

The probability under the null of observing a t statistic at least as unfavourable to the null as that actually observed is vanishingly slim. Reject the null that the coefficient on log parental income in the associated population linear regression ~~is~~ ~~is~~ at all reasonable levels of ~~con~~ significance.

$H_0: \beta_2 = 0$
$H_1: \beta_2 < 0$

t-statistic
$t = (\hat{\beta}_2 - \beta_2)/se(\hat{\beta}_2) = (-0.091 - 0)/0.003 = -30.333$

Under the null, given a sufficiently large iid random sample, by CLT, $t \overset{a}{\sim} N(0,1)$.

Reject the null if $t < -c_\alpha$, where c_α is the critical value for a one-sided test drawn from the standard normal distribution at level of significance $\alpha = 0.01$.

$P(N(0,1) < -c_\alpha) = 0.01 \Rightarrow c_\alpha = 2.326$

Reject the null. Conclude that the coefficient on ~~the~~ the rural dummy in the population linear regression of log child's income on log parental income and the rural dummy is negative. Holding ~~to~~ (log) parental income constant, children living in rural areas have lower income than children living in urban areas.

4ai $Y = E[Y|X] + \varepsilon$

$EY = E(E[Y|X] + \varepsilon)$

$= E(E[Y|X]) + E\varepsilon$

$= EY + E\varepsilon$

$\Rightarrow E\varepsilon = 0$

By linearity of expectation, law of iterated expectation

ii $\varepsilon = Y - E[Y|X]$

$E[\varepsilon|X] = E[(Y - E[Y|X])|X]$

$= E[Y|X] - E[E[Y|X]|X]$

$= E[Y|X] - E[Y|X]$

$= 0 = E\varepsilon$

$\Rightarrow \varepsilon$ is mean independent of X

By linearity of conditional expectation, conditioning

iii Let $f(X)$ be some arbitrary function of X. Suppose for reductio that ε is correlated with $f(X)$, i.e. $\mathrm{cov}(\varepsilon, f(X)) \neq 0$

$\mathrm{cov}(\varepsilon, f(X)) \neq 0$

$\Rightarrow E\varepsilon Ef(X) - E\varepsilon f(X) \neq 0$

$\Rightarrow E\varepsilon f(X) \neq 0$

$\Rightarrow E(E[\varepsilon f(X)|X]) \neq 0$

$\Rightarrow E(f(X)E[\varepsilon|X]) \neq 0$

$\Rightarrow E(f(X) \times 0) \neq 0$

$\Rightarrow E(0) \neq 0$

By reductio, ε is uncorrelated with $f(X)$. By generalisation, ε is uncorrelated with any function of X.

The above implications are by definition of covariance, results from (ai), (aii), LIE, conditioning.

b Variables X and Y are independent ($X \perp\!\!\!\perp Y$) iff their joint cumulative distribution function ($F_{XY}(x,y) := P\{X \leq x \wedge Y \leq y\}$) and their marginal cumulative functions ($F_X(x) := P\{X \leq x\}$, $F_Y(y) = P\{Y \leq y\}$) satisfy $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

Variable X is mean independent of Y iff $E[X|Y] = EX$, and similarly for Y.

Independence implies mean independence, but mean independence does not imply independence. For example, X and Y jointly distributed uniformly along the circumference of some circle in X,Y space are mean-independent but not independent.

Independence is reciprocal (evident from the definition) but mean independence is not. For example, X and Y jointly distributed uniformly along $y = x^2$ for $x \in [-1,1]$ are such that X is mean independent of Y but Y is not mean-independent of X.

c The conditional expectation of Y given X minimises the mean-squared prediction error of Y given only X. (or more generally the conditional expectation of Y given $\vec{X} = x_1, \ldots$ minimises the mean-squared prediction error of Y given only $\vec{X}$).

$E[Y|\vec{X}] = \mathrm{argmin}_{m(\cdot)} E(Y - m(\vec{X}))^2$

where m is some function of $\vec{X}$

$E(Y - E[Y|\vec{X}])^2$

$= E\varepsilon^2$

$= \mathrm{var}(\varepsilon)$

Consider arbitrary $m(\vec{X}) \neq E[Y|\vec{X}]$. Let $m(\vec{X}) = E[Y|\vec{X}] - f(\vec{X})$.

$E(Y - m(\vec{X}))^2$

$= E(\varepsilon + f(\vec{X}))^2$

$= \mathrm{var}(\varepsilon + f(\vec{X})) + [E(\varepsilon + f(\vec{X}))]^2$

$> \mathrm{var}(\varepsilon + f(\vec{X}))$

$= \mathrm{var}(\varepsilon) + \mathrm{var}(f(\vec{X}))$

$> \mathrm{var}(\varepsilon)$.

By results in (a)

So $E[Y|\vec{X}]$ minimises the mean-squared prediction error.

d Approximating the conditional expectation with a function that is linear in the variables is generally unrealistic. There is no reason to think that common ~~eco~~ relationships between economic variables are linear. For example, returns to ~~scale are generally decreasing because~~ ~~inves~~ scale are generally non-linear because of economies of scale. Approximation with a linear function is reasonable only for small, local effects.

Approximating the conditional expectation with a function that is linear in the parameters is more realistic because a suitably flexible model can be selected. This introduces the problem of model selection, which faces the bias-variance trade off.

Ultimately, linear models are common practice because they are estimable by common procedures like OLS.

Date

5a. $\beta_1 = cov(expeduc, girl)/var(girl)$

b β_1 captures the average causal effect of being a first born female on parental investments iff girl is exogenous, that is iff it is uncorrelated with the unmodelled determinants of expeduc collected in u, which is iff there is no omitted variable bias (~~omitte~~ simultaneity and measurement error in the independent variable are not plausible).

~~It is not~~ It is not reasonable to assume that there is no omitted variable bias. A birth sex ratio of 1.2 suggests some degree of sex-selective abortion is practiced. Such a practice is presumably voluntary ~~and/or elective~~ (by the parents) and so correlated with parental characteristics. One such characteristic is parental wealth and/or income. Presumably only ~~[illegible]~~ relatively wealthy and/or affluent parents can afford such sex-selective abortions. It seems that such sex-selective abortions "favour" boys, so being born female is likely to be negatively correlated with parental wealth and/or affluence. Plausibly, parental wealth is a positive ~~determina~~ causal determinant of parental investments ~~of~~ in education. Then, β_1 is likely to understate the average causal effect of being born female on parental investments in education.

This is clear in the omitted variable bias formula. ~~Let the fat~~

Consider the long population regression with both girl and ~~wealth~~ ~~affluence~~ wealth as regressors.

$$expeduc = \gamma_0 + \gamma_1 girl + \gamma_2 wealth + v$$

$$\beta_1 = cov(expeduc, girl)/var(girl)$$
$$= cov(\gamma_1 girl + \gamma_2 wealth, girl)/var(girl)$$
$cov(girl, v) = 0$ by construction
$$= \gamma_1 + \gamma_2 \, cov(wealth, girl)/var(girl)$$

By the above argument, $\gamma_2 > 0$, $cov(wealth, girl) < 0$, so $\beta_1 < \gamma_1$. Supposing that the long ~~p~~ population regression can be given a causal interpretation, β_1 understates the ~~causal effect~~ average causal effect of being born female on parental investment in education.

c $\gamma_1 = cov(expeduc, \tilde{girl})/var(\tilde{girl})$
(by the Frisch-Waugh-Lovell theorem)
where $\tilde{girl}$ is the residual of the auxiliary population regression of girl on nchild
$$girl = \delta_0 + \delta_1 nchild + \tilde{girl}$$
(where $E\,\tilde{girl} = 0$ and $cov(nchild, \tilde{girl}) = 0$). Intuitively, $\tilde{girl}$ is the component of girl whose variation is not (statistically, not necessarily causally) accounted for by variation in nchild. So γ_1 is the statistical effect of girl on expeduc, independent of nchild.

d No, nchild is an endogenous control. In a country where sex-selective abortion is practiced with a birth sex ratio of 1.2, such that sons are evidently preferred to daughters, it is plausible that being born female has a positive causal effect on nchild since parents are more likely to continue having children if their first born is a daughter, because they would like to have a son. The practice of including such endogenous controls is over-controlling.

Consider the causal models

$$expeduc = \beta_0 + \beta_1 girl + u$$
$$nchild = \delta_0 + \delta_1 girl + v$$

Date

6a On average, having an additional year of education is associated with having log wage higher by 0.08, holding age and male constant. ~~Equivalent~~ Equivalently, having an additional year of education is associated with having wage 1.0833 times as high as otherwise, holding age and male constant.

The required confidence interval is

$C = [\hat{\beta}_1 - c_\alpha se(\hat{\beta}_1), \hat{\beta}_1 + c_\alpha se(\hat{\beta}_1)]$

$= [0.08 - 2.576 \times 0.01, 0.08 + 2.576 \times 0.01]$

$= [0.05424, 0.10576]$

The random interval C contains the true value of β_1, which is the coefficient on education in the population linear regression of log wage on education, age, and male with 99% probability. At the 1% level of significance, fail to reject the null hypothesis of $\beta_1 = \beta_1^H$ for $\beta_1^H \in C$ against the alternative that $\beta_1 \neq \beta_1^H$.

b Regression (2) eliminates potential endogeneity in education due to omitted variable bias, in particular, due to the omission of determinants of log wages that are correlated with educ such as general cognitive ability and access to parent's social capital that are likely to be similar between identical twins.

$\hat{\gamma}_1$ has a higher value than $\hat{\beta}_1$. This suggests that the omitted variable bias in regression (1) eliminated by regression (2) is negative, which is the case iff education is positively correlated with negative determinants of log wages (risk aversion) or ~~that~~ negatively correlated with positive determinants of log wages.

$\hat{\gamma}_1$ has a higher standard error than $\hat{\beta}_1$. This suggests that the ~~variable~~ variance of residuals is greater for (2) than for (1) and/or that the variance of Δeducation is smaller than the variance of $\widetilde{\text{education}}$ (the component of education uncorrelated with age and male). The latter is likely because twins are likely to have similar levels of education, so the magnitude hence variance of Δ education is likely to be small.

Because variance in Δeducation is likely to be small, hence $se(\hat{\gamma}_1)$ is likely to be large, hypothesis tests on γ_1 are likely to be less powerful than hypothesis tests on β_1.

c $H_0: \delta_1 = 0$

$H_1: \delta_1 \neq 0$

t-statistic

$t = (\hat{\delta}_1 - 0)/se(\hat{\delta}_1) = 0.0025/0.0010 = 2.5$

Under the null, given a sufficiently large iid random sample, ~~t ≈ N(...)~~ by CLT, $t \overset{a}{\sim} N(0,1)$

Reject the null iff $|t| > c_\alpha$, where c_α is the critical value drawn from the standard normal distribution at the $\alpha = 0.05$ level of significance

$2\Phi(-c_\alpha) = \alpha = 0.05 \Rightarrow c_\alpha = 1.960$

Reject the null. Conclude that $\delta_1 \neq 0$, differences in birthweight explain differences in education.

Suppose that differences in birthweight are positively related to differences in log wages, independently of differences in education. Then, there is some causal determinant of log wages that is positively correlated with difference in birthweight hence positively correlated with differences in education, such as ~~poor physical~~ overall ~~or functional~~ physical health. ~~or functional physical fitness~~ Then, the estimate ~~of~~ in (2) overestimates the causal effect of education on log wages due to positive omitted variable bias.

Consider the long population regression of Δ log wage on Δeducation and Δx where x is the unobserved determinant.

~~$\Delta\varepsilon$~~ Δlog wages $= \pi_0 + \pi_1 \Delta\text{education} + \pi_2 \Delta x + \Delta v$

$\gamma_1 = cov(\Delta\text{log wage}, \Delta\text{education})/var(\Delta\text{education})$

$= cov(\pi_0 + \cdots + \Delta v, \Delta\text{education})/var(\Delta\text{education})$

$= cov(\pi_1 \Delta\text{education} + \pi_2 \Delta x, \Delta\text{education})/var(\text{education})$

$= \pi_1 + \pi_2 cov(\Delta x, \Delta\text{education})/var(\Delta\text{education})$.

Suppose that ~~to~~ the long regression ~~can be given~~ is causal, then the causal effect of Δeducation on Δ log wages is π_1. From the above, $\pi_2 > 0$ and $cov(\Delta x, \Delta\text{education}) > 0$, so $\gamma_1 > \pi_1$. $\therefore$ $\hat{\gamma}_1$ overestimates the causal effect of Δeducation on Δlog wages.

d Conduct a randomised control trial where some instrument which is either an offer or inducement to attend ~~an~~ additional education is randomly assigned, estimate the causal effect of ~~educat~~ education on log wages by ~~two-step~~ ~~2S~~ 2SLS or ICS regression.

~~A~~ A suitable instrument is relevant, i.e. correlated with education, exogenous, i.e. uncorrelated with the unmodelled determinants of ~~education~~ log wages, and excluded, ~~A~~ i.e. not itself a direct determinant of log wages. Exogeneity is guaranteed by successful random assignment of the instrument.

One plausible instrument is a small cash payment to parents in each year that their child remains in education over the duration of the study. If sufficiently large, this instrument is likely to be relevant, and if not too large, it is likely to be excluded.

2SLS and ICS would consistently estimate the local average treatment (education) effect, which is the sensitivity (to the instrument) weighted average ~~tees effect of educ~~ causal effect of education on log wages.

One limitation of this method is the LATE is likely to be different from the average causal effect in the population because low income families will likely be more responsive to the cash transfer and it is possible that additional ~~secon~~ ~~educ~~ education has a smaller effect on log wages for a low income child because of ~~a~~ ~~agate~~ other barriers to social mobility faced by low income children.