Quantitative Economics Paper 190531

1a $\text{var}(a+bX+cY)$

$=_1 E(a+bX+cY - E(a+bX+cY))^2$

$=_2 E(a+bX+cY-(a+bEX+cEY))^2$

$=_3 E(b(X-EX)+c(Y-EY))^2$

$=_4 E(b^2(X-EX)^2+c^2(Y-EY)^2+2bc(X-EX)(Y-EY))$

$=_5 b^2E(X-EX)^2+c^2E(Y-EY)^2+2bcE(X-EX)(Y-EY)$

$=_6 b^2\sigma_X^2+c^2\sigma_Y^2+2bc\,\sigma_{XY}$

where $=_1$ follows by definition of variance, $=_2$ by the given result (linearity of expectation), $=_3$ and $=_4$ by ~~alg~~ basic algebra, $=_5$ by linearity of expectation, and $=_6$ by definition of variance, covariance, σ_X^2, σ_Y^2 and σ_{XY}.

b $EZ = E(2X+3Y)$

$= 2EX + 3EY$

$= 14$

by substitution, linearity of expectation.

$\text{var}(Z) = \text{var}(2X+3Y)$

$= 4\text{var}(X) + 9\text{var}(Y)$

$= 100$

by substitution, result in (a), given $X \perp\!\!\!\perp Y$

$Z \sim N(14, \sqrt{100})$, by the result that any linear combination of normally distributed random variables is itself normally distributed.

c $P(Z>25)$

$= 1-P(Z\leq 25)$

$= 1-P(14+10Z' \leq 25)$

where $Z' \sim N(0,1)$

$= 1-P(Z' \leq 25-14/10)$

$= 1-P(Z' \leq 1.1)$

$= 1-0.8643$

from statistical tables

$= 0.1357$

2 Time series ~~random variable~~ $\{X_t\}$ does not Granger cause time series ~~random~~ ~~variable~~ $\{Y_t\}$ iff the mean squared forecast error of Y_{t+1} on the lags ~~of~~ y_t of Y_t and x_t of X_t is no less than the mean squared forecast error of Y_{t+1} on y_t alone. In symbols, this is iff

$$E(Y_{t+1}-E[Y_{t+1}|y_t,x_t])^2 = E(Y_{t+1}-E[Y_{t+1}|y_t])^2$$

This is iff $E[Y_{t+1}|y_t,x_t] = E[Y_{t+1}|y_t]$

$\{X_t\}$ Granger causes $\{Y_t\}$ iff the MSFE of Y_{t+1} on y_t and x_t is less than the MSFE of Y_{t+1} on y_t alone. In symbols this is iff

$$E(Y_{t+1}-E[Y_{t+1}|y_t,x_t])^2 < E(Y_{t+1}-E[Y_{t+1}|y_t])^2$$

A test of Granger causality ($\{X_t\}$ Granger causes $\{Y_t\}$) given the ADL (p, q) model

$Y_t = \alpha_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \sum_{i=1}^q \gamma_i X_{t-i} + u_t$

where $E[u_t | y_{t-1}, x_{t-1}] = 0$ by construction

is the following F test

$H_0: \forall i \in \{1,\dots,q\}: \gamma_i = 0$

$H_1: \not\forall i \in \{1,\dots,q\}: \gamma_i = 0$

~~Test compute the~~

Estimate the unrestricted model ~~given above by OLS~~.

$\hat{Y}_{t-1} = \hat{\alpha}_0 + \sum_{i=1}^p \hat{\beta}_i Y_{t-i} + \sum_{i=1}^q \hat{\gamma}_i X_{t-i} + \hat{u}_t$ by OLS.

Estimate the restricted model

$Y_{t-1} = \hat{\alpha}_0' + \sum_{i=1}^p \hat{\beta}_i' Y_{t-i} + \hat{u}_t'$ by OLS.

Compute the sum of squared residuals

$SSR = \sum_{i=1}^T \hat{u}_t$, $SSR' = \sum_{i=1}^T \hat{u}_i'$

Compute the F-statistic

$F = \frac{n-k-1}{q}\, SSR'-SSR/SSR$

where $n = T$, $k = p+q$

Under the null, supposing that $\{Y_t\}$ and $\{X_t\}$ are jointly stationary, the F statistic is distributed according to the usual $F_{q,\infty}$ distribution.

Reject the null iff $F > c_\alpha$, where c_α is the suitable critical value drawn from this distribution at the level of significance α.

Iff the null is rejected, conclude that $\{X_t\}$ Granger causes $\{Y_t\}$.

3a On average, a resident of Wales has log wage lower than a resident of England by 0.15, holding experience and gender constant. Equivalently, ~~a reside~~ on average, a resident of Wales has wage 0.86071 times that of a resident of England, holding experience and gender constant.

b On average, a person with two additional years of experience has a higher log wage by 2(0.05) = 0.10.

c The required confidence interval is

$$C = [2\times(0.05 - 1.645 \times 0.02), 2\times(0.05 + 1.645 \times 0.02)]$$
$$= \cancel{[0.0171, 0.0829]}\ [0.0342, 0.1658]$$

This random interval contains the true value of the effect of interest with 90% probability. ~~Any hypothesis test of~~

d $H_0: \beta_G = 0$
$H_1: \beta_G \neq 0$
where β_G is the coefficient on Gender in the given regression

t-statistic
$t = \hat{\beta}_G - 0 / se(\hat{\beta}_G) = 0.08/0.03 = 2.6667$

Under the null, given a ~~sufficiently~~ sufficiently large sample, supposing that it is an iid random sample, by CLT, $t \xrightarrow{d} Z \sim N(0,1)$

The p-value is the probability under the null of observing a test statistic as unfavourable to the null as that actually observed.
$p = P(|Z| > 2.6667)$
$= 2\Phi(-2.6667)$
$= 2(0.00383)$
from statistical tables
$= 0.00766$

~~Reject the null~~ Under the null, the probability of observing a test statistic as unfavourable to the null as that actually observed is 0.00766. ~~Reject~~ Reject the null at all levels of significance greater than 0.00766.

e Supposing that gender is uncorrelated with experience and the region dummies (which is plausible), there is no omitted variable bias, so the omission of these variables has no effect on the expected estimate of the coefficient on ~~of~~ gender. This estimate is expected to remain relatively unchanged.

~~Omission~~ These omitted regressors have non-zero coefficients in the "long" regression, so account for some of the ~~difference~~ variation in log wages. Then residuals in the "short" regression are larger and have higher variance. So the standard error of the ~~coefficient on~~ ~~β_G is~~ coefficient on gender in the short regression is larger.

1a $E\bar{Y}_n$ ~~$= E f ½$~~

$=_1 E(1/n \sum_{i=1}^{n} Y_i)$

$=_2 1/n \sum_{i=1}^{n} EY_i$

$=_3 1/n \sum_{i=1}^{n} \mu_Y$

$=_4 \mu_Y$

where $=_1$ follows by substitution, $=_2$ by linearity of expectation, $=_3$ ~~by the given fact~~ given that Y_i are iid and $EY_i = \mu_Y$ for all i, and $=_4$ by basic algebra.

$var(\bar{Y}_n)$

$=_1 var(1/n \sum_{i=1}^{n} Y_i)$

$=_2 1/n^2 \, var(\sum_{i=1}^{n} Y_i)$

$=_3 1/n^2 \sum_{i=1}^{n} var(Y_i)$

$=_4 1/n^2 \sum_{i=1}^{n} \sigma_Y^2$

$=_5 \sigma_Y^2/1000$

where $=_1$ follows by substitution, $=_2$ by the common result that $var(aX) = a^2 var(X)$, $=_3$ ~~by~~ given that Y_i are iid, $=_4$ by substitution, and $=_5$ by basic algebra.

Given that Y_i is iid, ~~by~~ and that the sample is large, by CLT, the ~~sample mean $\bar{Y}_n$ is approximately distributed according to N(~~ standardised sample mean $Z_n = \bar{Y}_n - \mu_Y / \sqrt{\sigma_Y^2/1000}$ is approximately distributed according to the standard normal distribution $Z \sim N(0,1)$. Then, the sample mean $\bar{Y}_n$ is approximately distributed according to $N(\mu_Y, \sqrt{\sigma_Y^2/1000})$.

This is independent of whether Y is normally distributed. The CLT applies regardless of the distribution of Y_i.

b $n = 400$, $\bar{Y}_n = 55$, ~~$s_{n,Y} = 10$~~ $s_{n,Y} = 10$

~~the $var(\bar{Y}_n)$~~

$se(\bar{Y}_n) = \sqrt{s_{n,Y}^2/n} = s_{n,Y}/\sqrt{n} = 10/20 = 1/2$

The required confidence interval is

C ~~$= [55 - 1.960 \times ½, 55 + 1.960 \times ½]$~~

$= [\bar{Y}_n - 1.960 \times se(\bar{Y}_n), \bar{Y}_n + 1.960 \times se(\bar{Y}_n)]$

$= [55 - 1.960 \times ½, 55 + 1.960 \times ½]$

$=$ ~~[53.9,~~ $[54.02, 55.98]$

The random interval C contains the true value of the population mean test score with 95% probability. Reject all hypothesised values outside this interval at the 5% level of significance.

c $H_0: \mu_Y = 50$

$H_1: \mu_Y > 50$

t-statistic

$t =$ ~~$\sqrt{n}$~~ $(\hat{\mu}_{n,Y} - 50)/se(\hat{\mu}_{n,Y})$

$= (\bar{Y}_n - 50)/se(\bar{Y}_n)$

$= (55 - 50)/s_{n,Y}/\sqrt{n}$

$= (55 - 50)/½$

$= 10$

Under the null, given a sufficiently large random sample, supposing that it is an iid random sample, by CLT, $t \xrightarrow{d} Z \sim N(0,1)$.

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

$p = P(Z > 10) = \Phi(-10) \approx 0$

from statistical tables.

The p-value is approximately zero. Under the null, the probability of observing a test statistic as unfavourable to the null as that actually observed is vanishingly small. Reject the null at all reasonable (including all conventional) levels of significance. Conclude that the mean test score of ~~out~~ students in Oxford is higher than the mean test score ~~of~~ in the UK.

$H_0: \mu_T - \mu_C = 0$

$H_1: \mu_T - \mu_C \neq 0$

~~$\hat{\mu}_T = 57$, $\hat{\mu}_C = 55$~~

$n_T = 900$, $\hat{\mu}_T = 57$, ~~$se(\hat{\mu}_T) = s_T/\sqrt{n_T} = 20/\sqrt{900} = 2/3$~~

$n_C = 400$, $\hat{\mu}_C = 55$, ~~$se(\hat{\mu}_C) = s_C/\sqrt{n_C} = 10/\sqrt{400} = ½$~~

$se(\hat{\mu}_T - \hat{\mu}_C) = \sqrt{s_T^2/n_T + s_C^2/n_C} = \sqrt{400/900 + 100/400}$

$= 0.83333$

Test statistic

$t = (\hat{\mu}_T - \hat{\mu}_C)/se(\hat{\mu}_T - \hat{\mu}_C)$

$= (57 - 55)/0.83333$

$= 2.4$

Under the null, given sufficiently large, independent, iid random samples, $t \xrightarrow{d} Z \sim N(0,1)$

Reject the null iff $t > c_\alpha$, where c_α is the critical value drawn from the $N(0,1)$ distribution at the $\alpha = 0.01$ level of significance.

$\alpha =$ ~~2Φ~~ $2\Phi(-c_\alpha) \Rightarrow c_\alpha = 2.576$

Fail to reject the null at the $\alpha = 0.01$ level of significance. Do not conclude that the populatio

Date

mean test score~~s~~ of treated students is significantly different from the ~~mean~~ population mean test score of untreated (control group) students.

5a $E[Y_i|X_i]$

$=_1 E[\beta_0 + \beta_{1i} X_i + u_i | X_i]$

$=_2 \beta_0 + E[\beta_{1i} X_i | X_i] + E[u_i | X_i]$

$=_3 \beta_0 + X_i E[\beta_{1i} | X_i] + E[u_i]$

$=_4 \beta_0 + X_i E[\beta_{1i}] + E[u_i]$

where $=_1$ follows by substitution, $=_2$ by linearity of conditional expectation ~~and by conditioning~~ and by conditioning, ~~likewise for $=_3$ and $=_4$~~ $=_3$ by conditioning and given u_i is mean independent of X_i, $=_4$ by conditioning and given β_{1i} is mean independent of X_i.

$E[Y_i|X_i]$ gives the ~~best~~ mean squared error-minimising prediction of Y_i given ~~X_i a~~ only X_i. ~~This f~~ From the above, $E[Y_i|X_i]$ is linear, so it is the mean squared error-minimising linear prediction of Y_i on X_i alone, i.e. it solves the population linear regression problem. Given that the population linear regression coefficients uniquely solve this problem, these coefficients ~~coeffi~~ coincide with the coefficients in $E[Y_i|X_i]$. In particular, the coefficient on X_i ~~coincides with~~ in the population linear regression coincides with ~~$E[\beta_{1i}]$, so~~ $E\beta_{1i}$, so $E\beta_{1i}$ is recoverable from such a regression.

b $E\beta_{1i}$ is the average causal effect of X_i on Y_i

c ~~By construction of the population linear regression model,~~

~~$\beta_{IV} = cov(Y, X^*)/var(X^*)$~~

β_{IV}

$=_1 cov(Y_i, X_i^*)/var(X_i^*)$

$=_2 cov(Y_i, \delta_0 + \delta_1 Z_i)/var(\delta_0 + \delta_1 Z_i)$

$=_3$ ~~$cov(Y_i,$~~ $\delta_1 cov(Y_i, Z_i)/\delta_1^2 var(Z_i)$

$=_4 cov(Y_i, Z_i)/\delta_1 var(Z_i)$

$=_5 cov(Y_i, Z_i)/var(Z_i) \times var(Z_i)/cov(X_i, Z_i)$

~~where $=_1$ follows by constr~~

$=_6 cov(Y_i, Z_i)/cov(X_i, Z_i)$

where $=_1$ follows by construction of the population linear regression model of Y_i on X_i^*, $=_2$ follows by substitution, $=_3$ by common results for cov, var, $=_4$ by basic algebra, $=_5$ by construction of the population linear regression of X on Z, $=_6$ by basic algebra.

β_{IV} coincides with that obtained by population ILS regression when there is only one instrument.

d ~~From the result of (c), $\delta_1 = E\pi_{1i}$ and β_{IV} $= E\gamma_{1i}$, where γ_{1i} is the coefficient~~

β_{IV}

$=_1 cov(Y_i, Z_i)/cov(X_i, Z_i)$

~~$= E(Y_i - EY_i)(Z_i - EZ_i) / E(X_i - EX_i)(Z_i - EZ_i)$~~

$=_2 cov(Y_i, Z_i)/var(Z_i) \;/\; cov(X_i, Z_i)/var(Z_i)$

$=_3 \frac{1}{E\pi_{1i}} \; cov(Y_i, Z_i)/var(Z_i)$

$=_4 \frac{1}{E\pi_{1i}} \; cov(\beta_0 + \beta_{1i} X_i + u_i, Z_i)/var(Z_i)$

$=_5 \frac{1}{E\pi_{1i}} \; cov(\beta_{1i} X_i, Z_i)/var(Z_i)$

$=_6 \frac{1}{E\pi_{1i}} \; cov(\beta_{1i} \pi_{1i} Z_i, Z_i)/var(Z_i)$

$=_7 \frac{1}{E\pi_{1i}} \; E(\beta_{1i}\pi_{1i} Z_i - E(\beta_{1i}\pi_{1i} Z_i))(Z_i - EZ_i)/E(Z_i - EZ_i)^2$

$=_8 \frac{1}{E\pi_{1i}} \; E(\beta_{1i}\pi_{1i})(Z_i - EZ_i)^2/E(Z_i - EZ_i)^2$

$=_9 \frac{1}{E\pi_{1i}} \; E(\beta_{1i}\pi_{1i}) E(Z_i - EZ_i)^2/E(Z_i - EZ_i)^2$

$=_{10} \frac{1}{E\pi_{1i}} \; E\beta_{1i}\pi_{1i}$

$=_{11} E\{\beta_{1i} \, \pi_{1i}/E\pi_{1i}\}$

where $=_1$ ~~follows by~~ follows by the result in (c), $=_2$ follows by basic algebra, $=_3$ follows by construction of the first stage regression and the result in (a), $=_4$ follows by substitution, $=_5$ by common results for cov, given u_i is independent of Z_i, $=_6$ by substitution, common results for cov, given V_i is independent of Z_i, $=_7$ by definition of cov and var, $=_8$ by linearity of expectation given β_{1i} and π_{1i} are independent of Z_i, likewise for $=_9$, $=_{10}$ by basic algebra, and $=_{11}$ by linearity of expectation

e β_{IV} is a local average causal effect. It gives ~~the~~ a weighted average causal effect, where the causal effect of each individual is weighted by that individual's ~~so~~ responsiveness to the instrument Z.

f $\pi_{1i} = \pi_1$ for all i. If this condition is satisfied, then each individual is equally ~~resp~~ responsive to the instrument, so the ~~weighte~~ weighted average coincides with the unweighted average, and the local average causal effect coincides with the average causal effect. $\pi_{1i} = E\pi_{1i}$ for all i, then $E\{\beta_{1i} \frac{\pi_{1i}}{E\pi_{1i}}\} = E\beta_{1i}$

β_{1i} and π_{1i} are ~~independent~~ uncorrelated. If this condition is satisfied, then variation in responsiveness to the instrument is uncorrelated with variation in ~~responsiveness~~ the causal effect, so ~~weighting the causal effect by responsiveness~~ the local average causal effect coincides with the average causal effect. $E\{\beta_{1i} \, \pi_{1i}/E\pi_{1i}\}$

$= \frac{1}{E\pi_{1i}} E\beta_{1i}\pi_{1i} = \frac{1}{E\pi_{1i}} E\beta_{1i} E\pi_{1i} = E\beta_{1i}$.

6ci 1997 test scores (S_{97}) are included in the regression as a proxy for relevant determinants of 1999 test scores (S_{99}) which should be controlled for. ~~Such omitted variables include~~.

~~1997 test sco S_{97} to~~ The coefficient on S_{97} is positive, and intuitively, we expect a student with higher S_{97} to have ~~hig~~ a higher S_{99}, so ~~we a priori and~~ we have a priori and a posteriori reason to think that ~~S~~ variation in S_{97} accounts for some of the variation in S_{99}. Then, including S_{99} in the regression ~~increases~~ decreases the magnitude and variance of the regression residuals, hence improves the precision of the estimator for the coefficient on offered voucher (OV). Then, hypothesis tests for the coefficient on OV in the population linear regression model are more powerful.

ii β_1 can be interpreted as the intention to treat effect, supposing that OV is successfully and unconditionally randomly assigned.

On average, a student who was awarded a voucher had a ~~S_{99} greate~~ higher S_{99} by 3.27, ~~but~~ holding S_{97} fixed. Supposing that OV is successfully and unconditionally ~~except for on S_{97}~~ randomly assigned, this can be interpreted as the indirect causal effect of OV on S_{99}.

This interpretation is plausible regardless of whether S_{97} is included as a regressor.

bi ~~The descriptive interpretation of δ_1 is as follows. On average.~~

~~δ_1 is the causal effect~~

δ_1 is the local average treatment (private school attendance P) effect (on S_{99}). On average, ~~of~~ the students induced to attend private school by ~~the vouch~~ OV, had a higher S_{99} than the students not OV by 4.41 exactly correct. average causal effect amg compliers

δ_1 ~~measures the~~ ~~estimates~~ is an estimate of the effect of P on S_{99} whereas β_1 is an estimate of the effect of OV on S_{99}.

ii That β_1 is smaller in magnitude than δ_1 is explained by ~~the fac~~ students not being entirely responsive to the instrument OV, i.e. not all students offered the voucher attended private school, so the intention to treat has a smaller (estimated) effect than actual treatment.

iii The causal effect of interest is the effect of private school attendance on test scores among households with children in grades 1-4 and with low incomes.

If the households whose children are already attending private school are excluded, then the instrument OV is not unconditionally randomly ~~assigne~~ assigned, but only conditionally (on prior private school attendance) ~~assigned~~ randomly assigned. Then, only the causal effect of private school attendance on test scores of students not already attending private school can be estimated.

For ~~such~~ students already attending private school, their responsiveness to the instrument is small or zero, so their inclusion in the sample has ~~a~~ only a small effect on the estimate of ~~a~~ the local average treatment effect, which is the responsiveness-weighted average treatment effect.

~~LATE~~

c The ~~local averag~~ estimate of local avege treatment effect is consistent only if the instrument OV is a valid instrument. A valid instrument is relevant, exogenous, and excluded.
☆ Int valid'y req also uncorr part^n & ret^n w under chars.
An instrument is relevant iff it is correlated with the endogenous variable that it is an instrument for. In this case, OV is a relevant instrument iff it is correlated with P. No direct regression of P on OV is reported from which it would be simple to verify relevance. There is strong a priori reason to think that OV is relevant (and also not a weak instrument) because the ~~magnitude~~ amount of the voucher is large and it is natural to suppose that parents respond to this incentive by being more likely to send their child to private school.

An instrument is exogenous iff it is uncorrelated with ~~unobserved~~ unmodelled determinants of the dependent variable. In this case, if OV is successfully randomly assigned, it is relevant. In this case, ~~OV is successfully randomly assigned conditional on~~ random assignment potentially fails because of non-participation and attrition. If the characteristics of households that remained in the study were not ~~identically~~ distributed identically to the households that did not participate or dropped out, then ~~the~~ OV is not randomly assigned within the population of interest. There is reason to think that ~~conditional ra~~ random assignment of the treatment fails because the households that remained in the study compared to those that dropped out were less ~~more~~ likely to be ~~black~~ black and more likely to be receiving welfare.

An instrument is excluded iff it is not itself a direct determinant of the dependent variable. This is clearly plausible in the present case.

d The results of this study are not likely to provide a reliable estimate of the effects of the given policy because of heterogeneity between the study population and the broader population and because of spillover effects, i.e. the assumption of individualistic treatment fails

The study population is students of low income households whereas the population of interest ~~in~~ for the proposed policy is ~~all child~~ students of all households. Plausibly, private schooling has a larger effect on academic performance for low income students because these students have less access to external opportunities and resources to.

There are potentially large spillover effects of the proposed policy, which could, for example, alter the composition of classes in public and private schools.

3a Given:

$\{u_t, v_t\}$ is iid

$Eu_t = Ev_t = 0$

$Eu_t^2 = \sigma_u^2$, $Ev_t^2 = \sigma_v^2$

$u_t \perp\!\!\!\perp v_t$

$y_t = \beta x_t + u_t$

$x_t = \gamma y_{t-1} + v_t$ for $t \geq 1$

$y_0 = 0$

$\beta\gamma = 1$

~~$(u_s, v_s) \perp\!\!\!\perp y_t$~~

$(u_t, v_t) \perp\!\!\!\perp (y_s, x_s)$ for all $s<t$

y_t

$= \beta x_t + u_t$

$= \beta(\gamma y_{t-1} + v_t) + u_t$

$= \beta\gamma y_{t-1} + \beta v_t + u_t$

$= y_{t-1} + \beta v_t + u_t$

Given that $\{u_t, v_t\}$ is iid, $\beta v_t + u_t$ is iid, so the required result obtains.

b Ey_t

$= E(y_{t-1} + \beta v_t + u_t)$

$= Ey_{t-1} + \beta E v_t + Eu_t$

$= Ey_{t-1}$

Then, by recursive substitution, $Ey_t = Ey_0 = 0$ for all t.

$\text{var}(y_t)$

$= \text{var}(y_{t-1} + \beta v_t + u_t)$

$= \text{var}(y_{t-1}) + \beta^2 \text{var}(v_t) + \text{var}(u_t)$

$= \text{var}(y_{t-1}) + \beta^2\sigma_v^2 + \sigma_u^2$

given that $(u_t, v_t) \perp\!\!\!\perp (x_s, y_s)$ for all $s<t$, $\{u_t, v_t\}$ is iid, and $u_t \perp\!\!\!\perp v_t$, by substitution and common results for var.

Then $\text{var}(y_t) = \text{var}(y_{t-1}) \Leftrightarrow \beta^2\sigma_v^2 + \sigma_u^2 = 0 \Rightarrow \sigma_u^2 = \sigma_v^2 = 0$

$\text{cov}(y_t, y_{t-h})$

$= \text{cov}(y_{t-h} + \sum_{s=t-h+1}^{t} (\beta v_s + u_s), y_{t-h})$

$= \text{cov}(y_{t-h}, y_{t-h})$

$= \text{var}(y_{t-h})$

given that $(u_t, v_t) \perp\!\!\!\perp (x_s, y_s)$ for all $s<t$, by common results for cov.

Then, $\text{cov}(y_t, y_{t-h})$ is time-invariant iff $\text{var}(y_t)$ is time-invariant which is iff $\text{var}(y_t)=0$ for all y_t, given that $\text{var}(y_0=0)=0$, which is iff $\sigma_u^2 = \sigma_v^2 = 0$, so $y_t, x_t, u_t, v_t = 0$ for all t.

~~$x_t = \gamma y_{t-1} + v_t$~~

$= \gamma y_{t-1} + v_t$

$= \gamma(\beta x_{t-1} + u_{t-1}) + v_t$

$= \gamma\beta x_{t-1} + \gamma u_{t-1} + v_t$

$= x_{t-1} + \gamma u_{t-1} + v_t$

Ex_t

$= Ex_{t-1}$

$\text{var}(x_t)$

$= \text{var}(x_{t-1}) + \gamma^2 \text{var}(u_{t-1}) + \text{var}(v_t)$

given that x_{t-1} is a function of x_{t-2}, u_{t-2}, and v_{t-1} and that u_{t-1} is independent of each of these.

$\text{var}(x_t) = \text{var}(x_{t-1}) \Leftrightarrow \sigma_u^2 = \sigma_v^2 = 0$

$\text{cov}(x_t, x_{t-h})$

$= \text{cov}(x_{t-h} + \sum_{s=t-h}^{t-1} (\gamma u_s + v_{s+1}), x_{t-h})$

$= \text{cov}(x_{t-h}, x_{t-h})$

$= \text{var}(x_{t-h})$

x_t is stationary iff Ex_t, $\text{var}(x_t)$ and $\text{cov}(x_t, x_{t-h})$ are time invariant which is iff $\text{var}(x_t)$ is time invariant which is iff ~~$u_t, v_t, x_t, y_t = 0$ for all t~~ $\sigma_u^2 = \sigma_v^2 = 0$ which is iff $y_t = x_t = u_t = v_t = 0$ for all t.

c y_t

$= y_{t-1} + \beta v_t + u_t$

$= y_{t-2} + \beta v_{t-1} + u_{t-1} + \beta v_t + u_t$

$\vdots$

$= y_0 + \sum_{s=1}^{t} (\beta v_t + u_t)$

$= \sum_{s=1}^{t} (\beta v_t + u_t)$

Given that $\{u_t, v_t\}$ is iid, $\beta v_t + u_t$ is iid, so y_t is a stochastic trend and has a stochastic trend component.

x_t

$= x_{t-1} + \gamma u_{t-1} + v_t$

$= x_{t-2} + \gamma u_{t-2} + v_{t-1} + \gamma u_{t-1} + v_t$

$\vdots$

$=$ ~~$x_0 + \sum_{s=1}^{t} (\gamma u_{s-1} + v_s)$~~

$x_1 + \sum_{s=2}^{t} (\gamma u_{s-1} + v_s)$

$= \gamma y_0 + v_1 + \sum_{s=2}^{t} (\gamma u_{s-1} + v_s)$

$=$ ~~$\sum_{s=1}^{t-1} (\gamma u_s + v_s) + v_t$~~ ~~$= v_1 + \Sigma$~~

$\sum_{s=1}^{t-1} (\gamma u_s + v_s) + v_t$

$\gamma u_s + v_s$ is iid, so $\sum_{s=1}^{t-1}$ is a stochastic trend component in x_t.

d conjecture that y_t and x_t are cointegrated with cointegrating coefficient $\theta = \beta$.

$$y_t - \beta x_t = u_t$$

Given that u_t is iid, u_t does not have a stochastic trend (otherwise independence fails). So y_t and x_t are cointegrated with cointegrating coefficient $\theta = \beta$

e $E[x_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_1 E[x_t + \gamma u_t + v_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_2 x_t + \gamma E[u_t \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_3 x_t + \gamma E[y_t - \beta x_t \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_4 x_t + \gamma(y_t - \beta x_t)$
$=_5 x_t + \gamma y_t - \gamma\beta x_t$
$=_6 \gamma y_t$

where $=_1$ follows by substitution of the result in (b), $=_2$ follows by linearity of conditional expectation, conditioning, given also that $v_{t+1} \perp\!\!\!\perp \mathcal{Y}_t, \mathcal{X}_t$ and $Ev_{t+1} = 0$, $=_3$ by substitution, $=_4$ by conditioning, $=_5$ by basic algebra and $=_6$ given $\gamma\beta = 1$.

$\cancel{E[x_{t+2} \mid \mathcal{Y}_t, \mathcal{X}_t]}$
$\cancel{= E[x_{t+1} + \gamma u_{t+1} + v_{t+2} \mid \mathcal{Y}_t, \mathcal{X}_t]}$
$\cancel{= \gamma y_t + \gamma E[u_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]}$
$\cancel{= \gamma y_t + \gamma E[y_{t+1} - \beta x_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]}$
$\cancel{= \gamma y_t + \gamma E[y_{t+1}]}$

$E[y_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_1 E[y_t + \beta v_{t+1} + u_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_2 y_t$

$=_1$ by substitution, $=_2$ by linearity of conditional expectation, conditioning, given that $v_{t+1}, u_{t+1} \perp\!\!\!\perp \mathcal{Y}_t, \mathcal{X}_t$ and $Ev_{t+1} = Eu_{t+1} = 0$

$\cancel{E[x_{t+2} \mid \mathcal{Y}_t, \mathcal{X}_t]}$
$\cancel{=_1 E[\gamma y_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]}$
$\cancel{=_2 \gamma y_t}$

~~$=_1$ by substitution of the~~

$E[x_{t+2} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_1 E[x_{t+1} + \gamma u_{t+1} + v_{t+2} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_2 E[x_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$=_3 \gamma y_t$

$=_1$ by substitution of the result in (b), $=_2$ by linearity of conditional expectation, independence, given $u_{t+1}, v_{t+2} \perp\!\!\!\perp \mathcal{Y}_t, \mathcal{X}_t$, $=_3$ by substitution of the result above.

Similarly

$E[x_{t+3} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$= E[x_{t+2} + \gamma u_{t+2} + v_{t+3} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$= E[x_{t+2} \mid \mathcal{Y}_t, \mathcal{X}_t]$
$= \gamma y_t$

By induction, $E[x_{t+h} \mid \mathcal{Y}_t, \mathcal{X}_t] = \gamma y_t$ for all $h \in \{1, 2, \dots\}$.

f $E[x_{t+1} \mid \mathcal{X}_t]$
$=_1 E[x_t + \gamma u_t + v_{t+1} \mid \mathcal{X}_t]$
$=_2 x_t + \gamma E[u_t \mid \mathcal{X}_t]$
$=_3 x_t + \gamma E u_t$
$=_4 x_t$

where $=_1$ follows by substitution, $=_2$ by linearity and independence given $v_{t+1} \perp\!\!\!\perp \mathcal{X}_t$ and $Ev_t = 0$, $=_3$ by independence given that $\mathcal{X}_t$ is a function of y_0, $\{u_s\}_0^{t-1}$ and $\{v_s\}_1^t$ hence $u_t \perp\!\!\!\perp \mathcal{X}_t$, $=_4$ given $Eu_t = 0$

$E(x_{t+1} - E[x_{t+1} \mid \mathcal{Y}_t, \mathcal{X}_t])^2$
$=_1 E(x_{t+1} - \gamma y_t)^2$
$=_2 E v_{t+1}^2$
$=_3 \sigma_v^2$

$=_1$ ~~by substitu~~ $=_2$ $=_3$ by substitution

$E(x_{t+1} - E[x_{t+1} \mid \mathcal{X}_t])^2$
$=_1 E(x_{t+1} - x_t)^2$
$=_2 E(\gamma u_t + v_{t+1})^2$
$=_3 \sigma_v^2 + \gamma^2 \sigma_u^2$

$=_1$ $=_2$ by substitution, $=_3$ by common results for var, given $u_t \perp\!\!\!\perp v_{t+1}$ hence $\mathrm{cov}(u_t, v_{t+1}) = 0$

The former prediction has a smaller MSFE and is more efficient. $\{y_t\}$ Granger causes $\{x_t\}$ because lags of the former carry useful information for prediction of the latter. This is evident simply from the equation $x_{t+1} = \gamma y_t + v_{t+1}$.

9a ip_t is approximately equal to the percentage change in IP_t ~~bet~~ in each period.

ip_t ~~= 100 × Δln IP_t~~
$= 100 \times \Delta \ln IP_t$
$= 100 (\ln IP_t - \ln IP_{t-1})$
$= 100 \ln (IP_t / IP_{t-1})$
$\approx 100 (IP_t / IP_{t-1} - 1)$

where the approximation $\approx$ is reasonable for small IP_t / IP_{t-1}

$\{ip_t\}$ appears to be mean stationary with a mean at or slightly above zero. $\{ip_t\}$ does not appear to be weakly stationary because of a structural break at or around 1961. In the epoch prior to this structural break, ip_t apparently has significantly higher variance than in the epoch subsequent to this structural break. Similarly there appears (although less clearly) to have been a structural break at or around 1985. $\{ip_t\}$ exhibits low persistence because excursions from the mean are short.

$\{O_t\}$ appears to be non-stationary because of an apparent structural break at or around 1972. In the epoch prior to this break, non-zero ~~observatio~~ observations of O_t were relatively infrequent and small compared to in the epoch subsequent to this break.

O_t is zero iff the percentage point difference between oil price at t and its maximum value in the past 12 months is ~~se~~ weakly negative, i.e. iff oil prices at t are no higher than at any point in the past 12 months. Many observations of O_t are zero because oil prices are relatively stable and do not frequently exceed 12 month highs. No observations are negative because O_t is the maximum of zero and the percentage point difference between oil prices in t and the 12 month high.

b Coefficients of the DL model can be given a causal interpretation iff each of the regressors is exogenous. ~~A regressor is~~ The sources of endogeneity are omitted variable bias, ~~measu~~ measurement error, and simultaneity.

Neither measurement error nor simultaneity are plausible in this context. Measurement error is not plausible in a well run study in which the data is correctly tabulated. Simultaneity is not plausible between O_{t-i} for $i \geq 1$ because it is not possible that present production has a causal effect on past oil prices. For ~~O_t~~, O_t ~~measurement error is~~ simultaneity is more plausible for a small producer whose production is likely to have a smaller feedback effect on oil prices.

Omitted variable bias is plausible because of such omitted variables as aggregate demand, which is positively correlated with oil prices and ~~is~~ a causal determinant of industrial production.

The assumption of exogeneity is more plausible for the UK because UK production is likely to have a smaller feedback effect on oil prices, ~~so iff that~~ then ~~it is more plausible that~~ simultaneity is less plausible.

c $H_0: \beta_0 = \cdots = \beta_{18} = 0$
$H_1: \beta_0 \neq 0$ or $\cdots$ or $\beta_{18} \neq 0$

F statistic is computed by $\frac{n-k-1}{q} \frac{SSR' - SSR}{SSR}$, where SSR' is the sum of squared residuals of the restricted model, SSR is the sum of squared residuals of the ~~re~~ unrestricted model, $n = T$ is the number of observations, $k = 19$ is the number of regressors, and $q = 19$ is the number of restrictions.

Under the null, supposing that $\{$~~O_t O_t~~ $O_t\}$ and $\{ip_t\}$ are jointly stationary, F is approximately distributed according to the $F_{q,\infty}$ distribution.

Reject the null iff $F > c_\alpha$, where c_α is the critical value drawn from the $F_{q,\infty}$ distribution at level of significance α.

Given that ~~the~~ the F-statistic has p-value 0.00, the probability under the null of observing a F statistic as unfavourable to the null as that actually observed is vanishingly small.

Reject the null at all conventional levels of significance, which are greater than $p = 0.00$.

Conclude that lags of O_t ~~gra~~ Granger cause ip_t. Supposing that the lags of O_t are exogenous, oil price shocks in the past 18 months have a causal effect on the growth rate of industrial production in the US.

ip_{t+1} decreases by 8.9.
ip_{t+2} decreases by 13.8.
And in general, the effect of $O_t=10$ and $O_{t+j}=0$ for $j>0$ on ip_{t+j} for $j=1,\dots,18$ is equal to $10\beta_j$ where β_j is the coefficient on ~~β_j O_j in the~~ O_{t-j} in the given regression.
ip_t increases by 2.3
ip_{t+3} ~~ip~~ decreases by 7.5

$\ln IP_{t+3}$
$= \ln IP_{t-1} + \Delta \ln IP_t + \Delta \ln IP_{t+1} + \Delta \ln IP_{t+2} + \Delta \ln IP_{t+3}$
$= \ln IP_{t-1} + 1/100(ip_t + ip_{t+1} + ip_{t+2} + ip_{t+3})$
$= \ln IP_{t-1} + -0.279$

The effect of $O_t=10$ and $O_{t+j}=0$ ~~(compared to~~ for $j>0$ (compared to $O_t=0$ and $O_{t+j}=0$) on $\ln IP_{t+3}$ is equal to -0.279

IP_{t+3} is $e^{-0.279}=0.75654$ times what it would have otherwise been (had there been no such shock at O_t).

e This model predicts that an oil shock has no effect on ~~production~~ growth rate of production. after two years, O_{t-24} is not a regressor in this number.

An ~~AR~~ ADL(p,q) model with $p+q<19$ could more parsimoniously capture dynamic causal effects over long horizons. An ADL(p,q) model has exponentially decaying dynamic causal effects in the infinite future.