Quantitative Economics Paper 210527

1a Given that, for all $i \in \{1, \dots, n\}$, $x_i \sim U[-1,1]$,

~~$f(x) = f$~~

Probability density function of x_i

$$f(x) = \begin{cases} 1/2 & \text{if } x \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$$

For even k,

$$\begin{aligned} Ex_i^k &= \int_{-\infty}^{\infty} x^k f(x)\, dx \\ &= \int_{-1}^{1} \tfrac{1}{2} x^k\, dx \\ &= \tfrac{1}{2}\left[\tfrac{1}{k+1} x^{k+1}\right]_{-1}^{1} \\ &= \tfrac{1}{2}\left[\tfrac{1}{k+1} - -\tfrac{1}{k+1}\right] \\ &= \tfrac{1}{k+1} \end{aligned}$$

For odd k,

$$\begin{aligned} Ex_i^k &= \int_{-\infty}^{\infty} x^k f(x)\, dx \\ &= \tfrac{1}{2}\left[\tfrac{1}{k+1} x^{k+1}\right]_{-1}^{1} \\ &= \tfrac{1}{2}\left[\tfrac{1}{k+1} - \tfrac{1}{k+1}\right] \\ &= 0 \end{aligned}$$

$$\Rightarrow Ex_i^k = \begin{cases} 1/(k+1) & \text{for even } k \\ 0 & \text{for odd } k \end{cases}$$

b For odd k,

$$\begin{aligned} \operatorname{var}(x_i^k) &= Ex_i^{2k} - (Ex_i^k)^2 \\ &= 1/(2k+1) - 0^2 \\ &= 1/(2k+1) \end{aligned}$$

c $n^{-1/2} \sum_{i=1}^n x_i^3 = n^{-1} \sum_{i=1}^n n^{1/2} x_i^3$

$E(n^{1/2} x_i^3) = n^{1/2} Ex_i^3 = 0$

$\operatorname{var}(n^{1/2} x_i^3) = n \operatorname{var}(x_i^3) = n/7$

$n^{-1/2} \sum_{i=1}^n x_i^3$ is the sample mean of $n^{1/2} x_i^3$

Given that $\{x_i\}$ is an iid random sample, by CLT, the standardised sample mean

$$\left(n^{-1/2} \sum_{i=1}^n x_i^3 - E(n^{1/2} x_i^3)\right) \Big/ \sqrt{\operatorname{var}(n^{1/2} x_i^3)/n}$$

converges in distribution to the standard normal distribution $N(0,1)$.

$$\left(n^{-1/2} \sum_{i=1}^n x_i^3\right) \Big/ \sqrt{1/7} \xrightarrow{d} N(0,1) \Rightarrow$$
$$n^{-1/2} \sum_{i=1}^n x_i^3 \xrightarrow{d} N(0, 1/7)$$

2a The given hypothesis test is an ~~test of~~ ADF test for a unit root in W_t. ~~The critic~~ The relevant hypotheses are

$H_0: \phi = 0$

$H_1: \phi < 0$

Under the null, the t statistics are ~~dist~~ distributed according to the DFcn distribution. Reject the null of a unit root if $t < -c_\alpha$, where c_α is the appropriate critical value drawn from the DFcn distribution at level of significance α. Suppose that the level of significance is $\alpha = 0.01$, then $c_\alpha = 3.43$.

~~From the table, compa~~

By comparing the reported t statistics against the critical value $-c_\alpha$, reject the null of a unit root in X_t, fail to do so for Y_t and Z_t. ~~X_t is a has a unit root~~ ~~cuts~~ Conclude that X_t is stationary, it has no unit root, and no stochastic trend. Conclude that each of Y_t and Z_t has a unit root, is non-stationary, and has a stochastic trend.

~~It could~~ Qualification: it could be the case, for each of Y_t and Z_t, that it has no unit root and has no stochastic trend because it ~~is~~ has a deterministic trend and is trend stationary. ~~is there~~, Trend stationary time series are better ~~accomm~~ accommodated under the null of this test, so we would fail to reject the null even if the variable has no unit root, because it is trend stationary. The Dickey-Fuller constant and trend test would then be appropriate.

b Spurious regression is the tendency to observe statistically significant relationships between time series that have order of integration 1. This is because such time series have a stochastic trend, and so tend to exhibit large swings of increase and decrease that can be matched to ~~the~~ swings in other such time series with surprising regularity.

From the above, X_t is stationary, it has order of integration zero, so spurious regression is not likely to be a problem in a regression of X_t on Y_t.

Consider the ADF test with $W_t = Y_t - Z_t$. ~~The~~ Compare the test statistic $t = -5.96$ against the critical value $-c_\alpha = -3.43$. Reject the null of a unit root in $Y_t - Z_t$. Conclude that $Y_t - Z_t$ is stationary, so Y_t and Z_t are cointegrated with cointegrating coefficient 1. Spurious regression is not a problem for a regression of Y_t on Z_t because the two variables are cointegrated, they share a stochastic (and deterministic, if there is one) trend. The relationship found between these two variables is genuine.

3ai [3]: $Y = \hat{\beta}_0 + \hat{\beta}_X X + \hat{\beta}_D D + \hat{\beta}_{XD} XD + \hat{u}$

$\partial Y / \partial X |_{D=1} = \hat{\beta}_X + \hat{\beta}_{XD} = 0.044 + 0.025 = 0.069$

Supposing that [3] can be interpreted causally, an increase in the daily disposable income of each household by £1 ~~per day~~ causes an increase in the daily food expenditure for a household with two children by £ 0.069.

ii Supposing that [3] can be interpreted causally, an increase in daily disposable income of each household by £1 causes an increase in the daily food expenditure for a household with one child by £ 0.044.

$\partial Y / \partial X |_{D=0} = \hat{\beta}_X = 0.044$

Then, the average change in daily food expenditure is $0.61 \times 0.069 + 0.39 \times 0.044 = 0.05925$ Average daily food expenditure increases by £ 0.05925 as a result.

$\bar{Y}_n = 1/n \sum_{i=1}^{n} Y_i$

$= 1/n \sum_{\{i: D_i = 1\}} Y_i + 1/n \sum_{\{i: D_i = 0\}} Y_i$

$= n_1/n \; 1/n_1 \sum_{\{i: D_i = 1\}} Y_i + n_0/n \; 1/n_0 \sum_{\{i: D_i = 0\}} Y_i$

$= n_1/n \, \bar{Y}_1 + n_0/n \, \bar{Y}_0$

$\partial \bar{Y}_n / \partial X = n_1/n \; \partial \bar{Y}_1 / \partial X + n_0/n \; \partial \bar{Y}_0 / \partial X$

$= n_1/n \; \partial Y / \partial X |_{D=1} + n_0/n \; \partial Y / \partial X |_{D=0}$

$= 0.61 \times 0.069 + 0.39 \times 0.044$

b $H_0: \beta_{XD} = 0$

$H_1: \beta_{XD} \neq 0$

t-statistic

$t = (\hat{\beta}_{XD} - 0) / se(\hat{\beta}_{XD})$

$= 0.025 / 0.022$

$= 1.1364$

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

Reject the null if $|t| > c_\alpha$, where c_α is the appropriate critical value drawn from the $N(0,1)$ distribution at level of significance $\alpha = 0.05$.

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

Fail to reject the null. Conclude that there is no statistically significant ~~relationship between~~ difference between (1) the relationship between food expenditure and income for households with one child, and (2) that for households with two children.

4a $P(W_{i,09} > 11)$
$= P(N(10, 3^2) > 11)$
$= P(N(0,1) > {11-10}/3)$
$= 1 - P(N(0,1) < 1/3)$
$= 1 - \Phi(1/3)$
$= 1 - 0.63057$
$= 0.36943$

For arbitrary worker i in S2, the probability that wage $W_{i,09}$ in last full time job before 31 DEC 2009 ~~ex~~ is greater than 11 is 0.36943.

b S1: supposing that the causal effect of interest is the causal effect of enrolment in the JTP (or, ~~equiva~~ that the causal effect is participation/completion and all persons enrolled participated/completed the ~~JT~~ JTP), ~~the causal effect of interest is consistently estimated by the difference in subsample means~~ given that the estimates can be ~~satu~~ interpreted causally, the causal effect of interest is consistently estimated by the difference in sample means. This estimate is $12.6 - 10.5 = 2.1$. The causal effect is an increase in wage by 2.1 \$/h.

S2: with the same qualifications as above, the causal effect of interest is consistently estimated by the difference in means before and after treatment. This estimate is $12.4 - 10 = 2.4$. The causal effect is an increase in wage by 2.4 \$/h.

c S1:

$H_0: \mu_{W,1} - \mu_{W,0} = 0$
$H_1: \mu_{W,1} - \mu_{W,0} \neq 0$
where $\mu_{W,1}$ is the population mean hourly wage (\$/h) in first full time job held after 1 JAN 2011 among those enrolled for the JTP, and $\mu_{W,0}$ is that among those not enrolled.

standard error
$se = \sqrt{s^2_{n_1,W_1}/n_1 + s^2_{n_0,W_0}/n_0}$
$= \sqrt{3.3^2/50 + 8.1^2/50}$
$= 0.64031$

t-statistic
$t = (\hat{\mu}_{W,1} - \hat{\mu}_{W,0} - 0)/se$
$= (12.6 - 10.5)/se$
$= 3.2796$

under the null, given a sufficiently large iid random sample, by CLT, $t \xrightarrow{d} N(0,1)$.

Reject the null iff $|t| > c_\alpha$, where c_α is the critical value drawn from the $N(0,1)$ distribution at level of significance $\alpha = 0.05$

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

Reject the null. Conclude that the JTP has a statistically significant effect on earnings.

S2:

$H_0: W_{i,11} - W_{i,09} = 0$
$H_1: W_{i,11} - W_{i,09} \neq 0$

$se = \sqrt{s^2_{n,\Delta D}/n}$
$= \sqrt{4.3^2/100}$
$= 0.43$

$t = (2.4 - 0)/0.43$
$= 5.5814$

under the null, given a sufficiently large iid random sample, $t \xrightarrow{d} N(0,1)$.

Reject the null iff $|t| > c_\alpha$, where c_α is the appropriate critical value drawn from the $N(0,1)$ distribution at the $\alpha = 0.05$ level of significance.

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

Reject the null. Conclude that the JTP has a statistically significant effect on earnings.

d Strength of S1: supposing that the 50 treated individuals were successfully randomly selected from among the ~~study~~ 100 study participants, assignment of treatment is uncorrelated with unmodelled determinants of wage, and so is exogenous. The difference in means (which coincides with the corresponding OLS regression coefficient) is then consistent for the causal effect of interest. Whether treatment is successfully randomly assigned can be tested using the ~~se~~ pre-treatment survey by a F test of the null hypothesis that coefficients on each regressor in the population ~~#~~ regression of treatment on survey responses is zero against the alternative that at least one is non-zero.

Weakness of S1: because the sample is split into treatment and control groups, the standard error on the ~~se~~ ~~tagpat~~ difference in means is ~~zero~~ larger than the standard error on the mean of the difference in S2. ~~so even if~~ so t-statistics for S1 are smaller in magnitude, S1 has less power than S2 to detect deviations from the null, of no difference in earnings due to the JTP.

Strength of S1: ~~St~~ Data from S1 allows for estimation of the counterfactual wage of ~~perso~~ individuals not enrolled in the JTP. This is simply $\hat{E}[W_{i,11} \mid T_i = 0]$, which is consistent for $E[W_{i,11} \mid T_i = 0]$. Data from S2 does not allow for such estimation because ~~as~~ ~~as~~ all individuals in S2 receive ~~treatment~~ treatment.

e The claim is dubious, and is not supported by the results of either ~~a~~ S1 or S2. Neither study is externally valid. ~~Exter The results of a study are externally valid iff the the the estimate of the causal effect of~~ A study is externally valid iff its results can be reliably extrapolated to the real world population or policy of interest. External validity fails if there is heterogeneity between the study population and the real world ~~popula~~ population of interest, between the intervention in the study and the policy of interest, between the surrogate outcome used in the study and the real outcome of interest, or if the assumption of individualistic treatment fails.

There is heterogeneity between the intervention in S1 and S2, namely enrolment in the JTP and the real world policy of interest, namely making the JTP available to the unemployed ~~are that~~ This is evident from the possibility that an unemployed worker could fail to enrol in the program once it is made available because ~~of~~ (1) he is not aware of the program, (2) he ~~does~~ is computer illiterate and cannot enrol online, (3) he is unmotivated to enrol. (possibly because forgoing a year of income to participate in training has a steep opportunity cost). If only a small subset of the real world population of unemployed workers enrol (supposing this is as good as random), ~~the~~ the effect of making the JTP available will be much smaller than the effect estimated by the studies.

The assumption of individualistic treatment is likely to fail because ~~wider~~ wide availability of the JTP in the population will likely increase labour supply hence reduce wage of any ~~for~~ worker who completes the JTP.

5a The state dummies are included as proxies to control for variation in such factors as quality of healthcare institutions, ~~demo~~ demographic characteristics, quality of nutrition across states. Such factors are causal determinants of bwght. ~~Omission of such factors for a suit~~ Failure to control for such variables introduces omitted variable bias if smoked is correlated with the state dummies, and ~~decreases th~~ increases the magnitude and variance of regression residuals, hence increases the standard error of ~~est~~ OLS estimators, decreasing their precision. Then, ~~infere~~ hypothesis tests on regression coefficients will be less powerful.

Only 49 of 50 state dummies are included because the 50 state dummies are perfectly multicollinear. ~~Inclusion of all 50 state dum dummies in the OLS regression~~ If all 50 state dummies are included, there is no unique solution to the OLS regression problem, the residuals of the auxiliary regressions (in FWL) are zero and ~~has~~ have zero variance, so the OLS estimators are undefined.

6 OLS estimation of (1) may fail to give a reliable estimate of the causal effect of smoked on bwght because of endogeneity in smoked, due to omitted variables. ~~bt~~

Smoked is likely to be correlated with other such determinants of bwght as whether ~~the~~ and how much the mother ~~drank~~ consumed alcohol during pregnancy, the quality of the mother's diet during pregnancy (correlation intermediated by income and wealth), and the quality of prenatal care received.

~~Long causal model~~ $\cancel{Y = \beta_0 + \beta_1 X}$
~~bwght~~ $\cancel{= \beta_0 + \beta_1}$ ~~smoked~~ $\cancel{+ \beta}$

This introduces omitted variable bias according to the omitted variable bias formula

$$\gamma_1 = \beta_1 + \beta_2 \pi_1$$

where γ_1 is the population regression coefficient on the endogenous regressor that ~~th~~ the OLS estimator is consistent for, β_1 is the causal effect of the endogenous regressor, β_2 is the causal effect of the omitted variable, and π_1 is the population regression coefficient on the endogenous regressor in the auxiliary regression of the omitted variable on the omitted regressor.

Smoked is likely to be negatively correlated with positive determinants of bwght (quality of diet, quality of prenatal care), and ~~negatively correlated with pos~~ positively correlated with negative determinants (drinking), so omitted variable bias is likely to be negative. OLS estimator of β_1 ~~OVB~~ underestimates the causal effect of interest (i.e. yields a more negative estimate).

c This is the strategy of instrumental variables. This strategy is viable iff the instrument is valid, i.e. relevant, exogenous, and excluded (and the controls are exogenous, ~~and~~ there is no perfect multicollinearity, the sample is iid, and there is finite kurtosis).

Relevance: tax is a relevant instrument for smoked iff ~~tax~~ smoked is correlated with tax (and is not a weak instrument if this correlation is large). An OLS regression of smoked on tax is reported in [1] and the coefficient on tax here is significant ($t = -7$), so there is empirical evidence for Relevance. Relevance is plausible given that ~~price~~ tax affects price hence consumption, intuitively.

Exogeneity: tax is an exogenous instrument iff it is uncorrelated with the unmodelled determinants of bwght collected in u. Exogeneity is plausible if we think that ~~the~~ tax is determined in a way that is uncorrelated with the other causal determinants of ~~bw~~ bwght, for example, if ~~the this~~ if it is determined by the relative strength of pro-tobacco and anti-tobacco lobbying groups. ~~Variance of taxes across states~~ Correlation of taxes with state-correlated determinants of bwght is not problematic because state controls are included. ~~Variation across correla of tax is unlikely to be correlated with the Empirica~~ The given data is silent on exogeneity. ~~Though~~

Exclusion: tax is excluded iff tax is not itself a direct causal determinant of bwght. This is obviously plausible.

d Supposing that tax is a valid instrument, the 2SLS estimate in [3] is consistent for the causal effect of interest (of smoked on bwght), whereas the OLS estimate (as explained earlier) ~~has~~ has omitted ~~variab~~ variable bias.

The standard error of the 2SLS estimator is much higher than that of the OLS ~~est~~ estimator because the component of smoked explained by tax, smoked* has ~~sign~~ much less variance than smoked. OLS minimises the variance of the residuals in the ~~s~~ sample analogue of the structural equation, so u_{OLS} has less variance than u_{2SLS}. ~~For ho~~ Assuming that residuals are ~~ho~~ homoskedastic,

$$se_{OLS} = n^{-1/2} \sqrt{\hat{var}(u_{OLS}) / \hat{var}(\tilde{X})} <$$
$$se_{2SLS} = n^{-1/2} \sqrt{\hat{var}(u_{2SLS}) / \hat{var}(\tilde{X}^*)}$$

ei ~~on~~ On average, a child's being conceived in a month where tax is higher by 1 dollar per pack had ~~so~~ higher bwght by 20 grams.

Supposing that [4] can be given a causal interpretation, which is iff tax is exogenous in [4], which is only if it is uncorrelated with unmodelled determinants of bwght that are not proxied for, ~~a US wide increase in cigarette~~ and supposing that the study is externally valid, which requires that ~~the~~ the study population is representative and that treatment is individualistic, the given policy increases ~~bwght by~~ average bwght by 20 grams.

ii Suppose that tax is a valid instrument, then [3] consistently estimates the causal effect of smoked on bwght in the study population, i.e. this study is internally valid.

The study is also externally valid if, in addition, the following are homogenous: (1) study population and real world population of interest, (2) study intervention and ~~p~~ real world policy of interest, (3) ~~sur~~ (potentially surrogate) outcomes measured in the study and real world outcomes of interest. External validity also requires that the ~~assump~~ assumption of individualistic treatment holds.

Then, the relevant qualifications here are that the public health program only has an effect on smoking rates and no other effect (direct or otherwise on bwght).

Then, the ~~tax~~ policy ~~tax~~ causes a ~~0.20~~ $-0.20 \times -564 = 112.8$ gram increase in average birthweight.

f The causal effect of interest can be estimated by ILS. This estimate is obtained by dividing the coefficient on ~~to~~ tax in [4] by the coefficient on tax in [1].

Neglecting controls,
structural equation: $bwght = \beta_0 + \beta_1 smoked + u$
first stage regression: $smoked = \pi_0 + \pi_1 tax + v$
By substitution,
reduced form regression:
$bwght = (\beta_0 + \beta_1 \pi_0) + \beta_1 \pi_1 tax + (u + \beta_1 v)$
By supposition of exogeneity of tax, construction of first stage regression, linearity of expectation and covariance, the reduced form regression is a population regression of bwght on tax. ~~So OLS~~

So OLS estimator in [4] is consistent for $\beta_1 \pi_1$, and that in [1] is consistent for π_1, so the corresponding quotient is consistent for $\beta_1 \pi_1 / \pi_1 = \beta_1$.

6a $Y = \beta_0 + \beta_1 X + u$

$u = \varepsilon(\gamma_0 + \gamma_1 X)$

$E\varepsilon = 0$, $E\varepsilon^2 = 1$, $\varepsilon \perp\!\!\!\perp X$

$X > 0$

$E[\varepsilon^2|X] = E\varepsilon^2 = 1$

where the first equality follows by independence, and the second is given.

b $E[u|X]$

$= E[\,\varepsilon(\gamma_0 + \gamma_1 X)|X]$

$= \gamma_0 E[\varepsilon|X] + \gamma_1 E[\varepsilon X|X]$

$= 0 + \gamma_1 X E[\varepsilon|X]$

$= 0$

by substitution, linearity of conditional expectation, ~~result from (a)~~, conditioning.

$E[\varepsilon|X] = E\varepsilon = 0$

$E[u^2|X]$

$= E[\,\varepsilon^2(\gamma_0 + \gamma_1 X)^2\,|X]$

$= E[\,\varepsilon^2(\gamma_0^2 + \gamma_1^2 X^2 + 2\gamma_0\gamma_1 X)|X]$

$= \gamma_0^2 E[\varepsilon^2|X] + \gamma_1^2 E[\varepsilon^2 X^2|X] + 2\gamma_0\gamma_1 E[\varepsilon^2 X|X]$

$= \gamma_0^2 + \gamma_1^2 X^2 + 2\gamma_0\gamma_1 X$

by substitution, algebraic expansion, linearity of conditional expectation, result in (a), conditioning.

~~var(ui|xi)~~

~~= E(ui~~

$var(u|X)$

$= E[\,(u - E[u|X])^2|X]$

$= E[\,u^2 + (E[u|X])^2 - 2uE[u|X]\,|X]$

$= E[u^2|X] + E[\,(E[u|X])^2|X] - 2E[uE[u|X]|X]$

$= E[u^2|X] + E[0^2|X] - 2E[0|X]$

$= E[u^2|X]$

$= \gamma_0^2 + \gamma_1^2 X^2 + 2\gamma_0\gamma_1 X$

$= (\gamma_0 + \gamma_1 X)^2$

by substitution, algebraic expansion, linearity of conditional expectation, earlier results, algebraic factorisation.

c $Eu = E(E[u|X]) = E0 = 0$

by law of iterated expectations, result in (b).

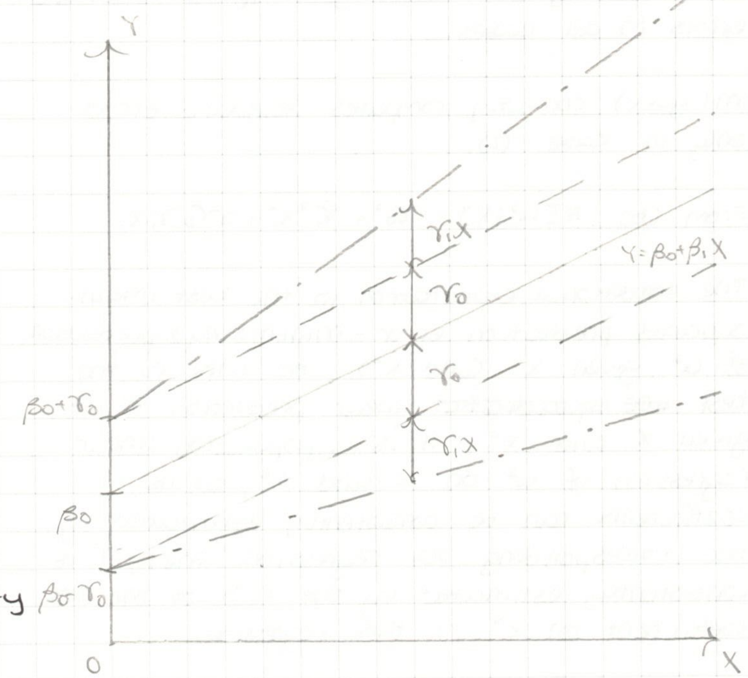
$EXu = E(E[Xu|X]) = E(XE[u|X]) = E0 = 0$

by law of iterated expectations, conditioning, result in (b).

orthogonality holds in the model $Y = \beta_0 + \beta_1 X + u$, i.e. $Eu = EXu = 0$, so this model is a population linear regression model, so its parameters can be consistently estimated by OLS. This consistency follows from the consistency of $\widehat{cov}$, $\widehat{var}$, $\hat{E}$ for their population counterparts. so $\hat\beta_1 = \widehat{cov}(Y,X)/\widehat{var}(X) \xrightarrow{p} cov(Y,X)/var(Y,X) = \beta_1$ and $\hat\beta_0 = \hat{E}Y - \hat\beta \hat{E}X \xrightarrow{p} EY - \beta_1 EX = \beta_0$.

d i $\gamma_0 > 0$, $\beta_1 > 0$

$sd(u|X) = \sqrt{var(u|X)} = \gamma_0 + \gamma_1 X$



For $\gamma_1 = 0$, Y ~~and X are distributed~~ is ~~that~~ distributed with constant variance and standard deviation around the line $Y = \beta_0 + \beta_1 X$. The dashed lines contain ~ 68% of the observations.

For $\gamma_1 > 0$, Y is distributed with increasing variance and standard deviation around the line $Y = \beta_0 + \beta_1 X$. The dashed and dotted lines contain ~ ~~63~~ 68% of observations.

e clothing is a necessity. Low income ~~house~~ households' expenditure on clothing is driven largely by need (to replace worn or outgrown clothes) and is constrained by income, so there is less variance in clothing expenditure among low income households. High income households' expenditure on clothing is driven by need but also by such factors as taste and conspicuous consumption, which vary more across households. their expenditure on clothing is also less constrained.

f lm_robust(y~x) computes heteroskedasticity robust standard errors, lm(y~x) computes standard errors assuming homoskedasticity. Residuals are homoskedastic iff $E[u^2|X] = Eu^2$.

$E[u^2 | X] = Eu^2 \iff$

$\cancel{(\gamma_0 + \gamma_1 X)^2 =}$

$\gamma_0^2 + \gamma_1^2 X^2 + 2\gamma_0\gamma_1 X = \gamma_0^2 + \gamma_1^2 EX^2 + 2\gamma_0\gamma_1 EX \iff$

$\gamma_1^2 X^2 + \gamma_0\gamma_1 X = \gamma_1^2 EX^2 + 2\gamma_0\gamma_1 EX \iff$

$\gamma_1 = 0$ or X is a constant (reject).

lm_robust (y~x) correctly computes standard errors in all cases.

lm(y~x) correctly computes standard errors only in case (i).

g From (b), $E[u^2 | X] = \gamma_0^2 + \gamma_1^2 X^2 + 2\gamma_0\gamma_1 X$.

The conditional expectation is the best (mean-squared prediction error - minimising) prediction of u^2 given X (and X^2), so this is the best ~~the quadratic~~ linear prediction of u^2 given X and X^2, it is a population linear regression of u^2 on X and X^2, so its coefficients can be consistently estimated by the corresponding OLS regression. Then, γ_1 is consistently estimated by the root of the OLS coefficient on X^2 in this regression.

OLS regression of Y on X consistently estimates β_0, β_1 (by argument in (c)). Then the OLS residuals $\hat{u} = Y - \hat{\beta}_0 - \hat{\beta}_1 X$ consistently estimate u. Then the above procedure can be performed using $\hat{u}$ rather than u. ~~The~~ γ_1 is consistently estimated by the ~~at~~ root of the coefficient on X^2 in the OLS regression of $\hat{u}$ on X, X^2.

9a We would be concerned about the ~~poss~~ possibility of ~~a~~ spurious regression in [1], [2], ~~[4]~~ [5], [6] and [7]. Given that Y_t and X_t are both $I(1)$, so are Y_{t-1}, Y_{t-2}, X_{t-1}, and X_{t-2}.

Spurious regression is the tendency to find statistically significant relationships between ~~indepe~~ time series ~~that~~ that have order of integration 1 that are in fact independent. This occurs because such time series have a stochastic trend, and exhibit large swings of increase and decrease that can be matched to swings in ~~the~~ the other series with surprising regularity.

Regression of Y_t on its own lags, even though both ~~have order~~ are $I(1)$ does not pose the problem of spurious regression because, presumably, the relationship between Y_t and its lags is genuine. This is implied if Y_t follows some AR(p) $p \geq 1$ model or some ADL (p, q) $p \geq 1$ model.

The problematic regressions regress Y_t on one or ~~potenti~~ more lags of X_t, which are potentially independent $I(1)$ time series variables.

b Consider model [1]. Given that both Y_t and X_t are $I(1)$, spurious regression is a potential problem. OLS regression of Y_t on X_t ~~gives~~ yields an estimate of a genuine rather than spurious relationship between Y_t and X_t iff Y_t and X_t are cointegrated. This is iff there exists some cointegrating coefficient θ such that equilibrium error $\varepsilon_t = Y_t - \theta X_t$ is stationary.

Suppose that no hypothesised value of θ exists, then estimate θ by ~~a~~ OLS regression of Y_t on X_t. Then, compute the estimated equilibrium errors $\hat{\varepsilon}_t = Y_t - \hat{\theta} X_t$. Perform ADF tests for a unit root in $\hat{\varepsilon}_t$ for a suitable range of ps. ~~If the null of a unit root~~ compare the t-statistics from such tests against suitably adjusted Engle-Granger critical values. If the null of a unit root is rejected, conclude that $\hat{\varepsilon}_t$ is stationary and that Y_t and X_t are cointegrated with cointegrating coefficient $\hat{\theta}$. Otherwise, conclude that Y_t and X_t are not cointegrated, and any relationship found between the two is potentially spurious.

c [1] would consistently estimate their cointegrating relationship. Suppose that Y_t and X_t are cointegrated with cointegrating coefficient θ. Given that the two variables are $I(1)$, each has some stochastic trend component. Given that the two are cointegrated with cointegrating coefficient θ, this stochastic trend component is common to both Y_t and X_t (as is any deterministic trend component with θ) ~~and and has factor~~. This stochastic trend component (and any deterministic trend component) is eliminated in $\varepsilon_t = Y_t - \theta X_t$ (because otherwise ε_t would have a trend component and would not be stationary).

OLS estimator $\hat{\beta}_1$ (and $\hat{\beta}_0$) in OLS regression $Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{u}_t$ minimise mean-squared prediction error $E\hat{u}_t^2$. For $\hat{\beta}_1 \neq \theta$, $\hat{u}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$ has a stochastic trend, whose variance ~~increases~~ is large and increasing in t, so as t becomes large, $E\hat{u}_t^2$ becomes large. ~~As~~ For $\hat{\beta}_{01} = \theta$, ~~$Y_t - \hat{\beta}_0 - \beta$~~ $\hat{u}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t = \varepsilon_t - \hat{\beta}_0$ is stationary and has fixed variance. So as t becomes large, $\hat{\beta}_1$ converges in probability to θ. Elimination of the common stochastic trend is necessary to minimise mean-squared prediction error. OLS regression of Y_t on X_t consistently estimates their cointegrating relationship.

The same argument cannot be given for Y_t and X_{t-1} unless the common stochastic trend is ~~a~~ a cumulation of stationary shocks, then Y_t and X_{t-1} are cointegrated with cointegrating coefficient θ.

d Only models [6] and [7] can be used to investigate Granger causality. $\{X_t\}$ Granger causes $\{Y_t\}$ iff lags of X_t contain useful information for the prediction of Y_t ~~inde~~ independently of the lags of Y_t. Formally, this is iff

$$E(Y_{T+1} - E[Y_{T+1} | \mathcal{Y}_T])^2 = E(Y_{T+1} - E[Y_{T+1} | \mathcal{Y}_T, \mathcal{X}_T])^2$$

$$\Leftrightarrow$$

$$E[Y_{T+1} | \mathcal{Y}_T] = E[Y_{T+1} | \mathcal{Y}_T, \mathcal{X}_T]$$

where $\mathcal{Y}_T = Y_T, Y_{T-1}, \ldots$ and $\mathcal{X}_T = X_T, X_{T-1}, \ldots$

The test of Granger causality ($\{X_t\}$ Granger causing $\{Y_t\}$), given some ADL (p, q) model ~~that reg~~ (p=q) that regresses Y_t on p lags of Y_t and q=p lags of X_t is the F (or t) test of the hypothesis that each coefficient on a lag of X_t is zero. So only models [6] and [7] are appropriate.

e The relevant test is possible for both [6] and [7].

[6]:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

t-statistic

$t_\alpha = (\hat{\beta}_1 - 0)/se(\hat{\beta}_1) = 0.19/0.07 = 2.7143$

Under the null, given that $\{Y_t\}$ and $\{X_t\}$ are (jointly) stationary, by CLT, $t \xrightarrow{d} N(0,1)$.

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

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

Reject the null, conclude that $\{X_t\}$ Granger causes $\{Y_t\}$.

[7]:

$H_0: \beta_1 = \beta_2 = 0$

$H_1: \beta_1 \neq 0$ or $\beta_2 \neq 0$

~~F statistic~~

$R^2_{un} = 0.20 \Rightarrow SSR_{un}/TSS = 1 - 0.20 = 0.80$

$R^2_{rs} = 0.14 \Rightarrow SSR_{rs}/TSS = 1 - 0.14 = 0.86$ (from [4])

$SSR_{rs} - SSR_{un}/SSR_{un} = 0.86 - 0.80/0.80 = 0.075$

$F = T-k-1/q \; SSR_{rs} - SSR_{un}/SSR_{un} = 100-4-1/2 \, (0.075)$

$= 3.5625$

Under the null, given that $\{Y_t\}$ and $\{X_t\}$ are jointly stationary, $F \xrightarrow{d} F_{q,\infty} = F_{2,\infty}$

Reject the null iff $F > c_\alpha$, where c_α is the ~~suitable or~~ appropriate critical value drawn from the $F_{2,\infty}$ distribution at level of significance $\alpha = 0.05$. $c_\alpha = 3.00$

Reject the null. Conclude that $\{X_t\}$ Granger causes $\{Y_t\}$.

f [1] and [5] are simply inappropriate because X_{t+1} is not known (or not ~~be~~ available for forecasting Y_{t+1}).

It is not appropriate to simply select the model that minimises R^2 because a larger model always has a lower R^2. But a larger model may overfit the data, i.e. be too ~~comp~~ complex or flexible relative to the real statistical process that generates the observations. An overfit model will be too sensitive to noise ~~and~~ in the data, and ~~will~~ thus will yield worse (higher mean-squared prediction error) predictions on out-of-sample data.

It is more appropriate to select a model using one of the information criterion. For large T, the ~~model~~ BIC tends to select the model with exactly the true number of lags, ~~(this does not necessarily~~ so we use BIC here.

$BIC = \ln SSR_m/T + m \ln T/T$

Of the eligible (not [1], [5]) models, BIC selects [6]. Use [6] for forecasting.