Quantitative Economics Paper 220527

1ai $P\{X \in [0, 0.5]\} = 1 \Rightarrow \mathcal{X} = [0, 0.5]$

$P\{Y \leq y \mid X = x\} = \begin{cases} 0 & \text{if } y < 0 \\ 1 - \gamma x & \text{if } y \in [0, 1) \\ 1 & \text{if } y \geq 1 \end{cases}$

for $x \in \mathcal{X}$, ~~for~~ where $\gamma \in [0, 2]$

$f_{Y|X}(y \mid 0.5) = \begin{cases} 1 - \gamma x = 1 - 0.5\gamma & \text{for } y = 0 \\ \gamma x = 0.5\gamma & \text{for } y = 1 \\ 0 & \text{for all other } y \end{cases}$

ii $f_{Y|X}(y \mid x) = \begin{cases} 1 - \gamma x & \text{for } y = 0 \\ \gamma x & \text{for } y = 1 \\ 0 & \text{for all other } y \end{cases}$

b $E[Y \mid X = x] = \sum_{y \in \mathcal{Y}} y f_{Y|X}(y \mid x)$
$= 0(1 - \gamma x) + 1\gamma x$
$= \gamma x$

c $E[Y \mid X] = \gamma X$
$EY = E(E[Y \mid X]) = E(\gamma X) = \gamma EX = 0.25\gamma$
By definition of conditional expectation, law of iterated expectations, linearity of expectation, and the result that $EX = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{0.5} 2x\, dx = 0.25$.

d From the above, $EY = \gamma EX$. By the consistency of $\hat{E}X := 1/n \sum_{i=1}^{n} x_i$ for EX (in an iid random sample), and the same property for Y, $\hat{\gamma} = \hat{E}Y / \hat{E}X$ is consistent for $\gamma = EY/EX$.

2a $Y_t = \beta_0 + \sum_{i=1}^{p_0} \beta_i Y_{t-i} + u_t$
$E[Y_{T+1} \mid \mathcal{Y}_T]$
$= E[\beta_0 + \sum_{i=1}^{p_0} \beta_i Y_{T+1-i} + u_{T+1} \mid \mathcal{Y}_T]$
$= \beta_0 + \sum_{i=1}^{p_0} \beta_i E[Y_{T+1-i} \mid \mathcal{Y}_T] + E[u_{T+1} \mid \mathcal{Y}_T]$
$= \beta_0 + \sum_{i=1}^{p_0} \beta_i Y_{T+1-i}$
By substitution, linearity of conditional expectation conditioning, given that $E[u_t \mid \mathcal{Y}_{t-1}] = 0$ for all t.

$E[Y_{T+1} \mid \mathcal{Y}_T]$ is the optimal (mean-squared forecast error-minimising) forecast of Y_{T+1} given $\mathcal{Y}_T = Y_T, Y_{T-1}, \ldots$. From the above, this is an AR(p_0) model.

bi $\{\varepsilon_t\}$ is an iid sequence, so $\{X_t\}$ is an iid sequence, and it is weakly stationary.

ii $EX_t = (-1)^t E\varepsilon_t = (-1)^t \mu$
EX_t is time-invariant iff $\mu = 0$
$var(X_t) = var((-1)^t \varepsilon_t) = (-1)^{2t} var(\varepsilon_t) = \sigma^2$
$var(X_t)$ is time-invariant ~~iff~~ for all μ, σ^2
$cov(X_t, X_{t-h})$
$= cov((-1)^t \varepsilon_t, (-1)^{t-h} \varepsilon_{t-h})$
$= (-1)^{2t-h} cov(\varepsilon_t, \varepsilon_{t-h})$
$= 0$
by linearity of expectation, given that $\varepsilon_t, \varepsilon_{t-h}$ are independent for $h \neq 0$, so $cov(X_t, X_{t-h})$ is time-invariant for all μ, σ^2.
X_t is stationary iff $\mu = 0$

iii $X_t = \rho X_{t-1} + \varepsilon_t$ with $X_1 = \varepsilon_1$
EX_t
$= \rho EX_{t-1} + E\varepsilon_t$
$= \rho EX_{t-1} + \mu$
$EX_t = EX_{t-1} \Leftrightarrow$
$EX_t = \rho EX_t + \mu \Leftrightarrow$
$EX_t = \mu / 1 - \rho$
$var(X_t)$
$= \rho^2 var(X_{t-1}) + \sigma^2$
$var(X_t) = var(X_{t-1}) \Leftrightarrow$
$var(X_t) = \sigma^2 / 1 - \rho^2$
$cov(X_t, X_{t-1}) = \rho\, var(X_{t-1})$
$cov(X_t, X_{t-h}) = \rho^h var(X_{t-h})$
given that each X_t is a function of $\{\varepsilon_s\}_{s=1}^{t}$ and $\{\varepsilon_t\}_{t=1}^{T}$ is iid.
X_t is ~~iid~~ stationary only if $EX_t = \mu / 1 - \rho = EX_1 = E\varepsilon_1 = \mu$
$\Leftrightarrow \rho = 0$ or $\mu = 0$.

If $\rho = 0$, then $X_t = \varepsilon_t$, which is iid hence trivially stationary.

If $\mu = 0$, then X_t is stationary iff, additionally, $var(X_t)$ is time-invariant (which implies that $cov(X_t, X_{t-h})$ is time-invariant). This is iff $\rho \in (-1, 1)$ such that $var(X_t) \geq 0$ and is defined, and $var(X_t) = \sigma^2 / 1 - \rho^2 = var(\varepsilon_1) = \sigma^2$, so $\rho = 0$.
X_t is stationary iff $\rho = 0$.

3a i No. degree is not plausibly exogenous because of omitted variables. Degree is correlated with unmodelled determinants of wage. For example, cognitive ability is plausibly a positive determinant of wage, and persons with higher cognitive ability are more likely to complete a degree programme and earn a degree. So the OLS coefficient on degree in the OLS regression of wage on degree will suffer from positive omitted variable bias.

$\hat{\gamma}_1 \xrightarrow{p} \gamma_1 = \beta_1 + \beta_2 \, cov(X_2, X_1)/var(X_1)$

where $\hat{\gamma}_1$ is this OLS estimator, γ_1 is the population regression coefficient that it is consistent for, β_1 is the causal effect of interest, β_2 is the causal effect of such unmodelled determinants ~~of~~ as cognitive ability on wage.

ii 2SLS regression of wage on degree with grsch as an instrument consistently estimates the causal effects of interest iff grsch is a valid instrument, which is iff grsch is relevant, exogenous, and excluded.

Relevance is plausible. A person who graduated in 1992 is more likely to have a degree than a person who graduated in 1991 because there simply were more ~~a~~ universities. Summary statistics in the given table support Relevance: the group mean of degree is higher among 1992 graduates than among 1991 graduates.

~~Exclusion is plausible iff if reasonable if we assume that wage inflation is relatively low, such that a later graduation in itself and a later entry into the workforce~~

Exclusion is plausible, ~~a~~ if employers do not discriminate on the basis of an employee's year of graduation in setting wages. ~~and~~ This is not entirely reasonable. ~~Presumably an increase a sharp increase in the proportion of degree holders~~ Employers may respond to the ~~pass~~ policy change by lowering wages.

Exogeneity is not plausible. grsch is likely to be strongly correlated with such determinants of wage as the ~~economic~~ macroeconomic conditions at the time one entered the workforce, and ~~are~~ the year one turned 30 (which has an effect on wage due to wage inflation).

So 2SLS will not yield an entirely reliable estimate either.

b In the case of a binary treatment, the OLS estimator of the population regression coefficient is equal to the difference in sample means.

The coefficient on grsch in the reduced form regression of wage on grsch is
2.351 - 2.347 = 0.004

The coefficient on grsch in the first stage regression of degree on grsch is
0.281 - 0.229 = 0.052

By ILS, the causal effect of interest is estimated by 0.004 / 0.052 = 0.076923

~~Obtaining~~ Having a degree causes an increase in log wages at age 30 by 0.76923. This is equivalent to an $e^{0.76923} = 1.0800$ fold increase in wage at age 30.

c The causal effect of degree on wage is likely to be heterogenous because, for example, different occupations attach different valuations to having a degree.

2SLS regression in (b), supposing that grsch is a valid instrument, consistently estimates the local average treatment effect, which is the responsiveness-weighted average causal effect of degree on wage. ~~Students~~ Individuals who graduated from former polytechnics would have high responsiveness (to the instrument grsch), so ~~resp~~ this weighted average is not likely to coincide with the unweighted average treatment effect consistently estimated by OLS in (c). Even if the ~~populat~~ samples in both studies are large, their estimators are consistent for different quantities, and so will not coincide.

d No. External validity is not plausible for such an intervention because of significant heterogeneities between the 1991 ~~and~~ / 1992 graduating cohorts and the 2023 graduating cohort. The different cohorts will be entering economies that reward different sorts of skills, and that are to different degrees open to ~~labour to~~ the world economy.

4a Causal model $Y = \beta_0 + \beta_1 X + u$
Population regression $Y = \delta_0 + \delta_1 X + e$,
where $Ee = cov(X, e) = 0$ by construction.

If $Eu = cov(X, u) = 0$, then the causal model coincides with the population regression because the population regression ~~para~~ coefficients are ~~the~~ uniquely such that $Ee = cov(X, e) = 0$.

~~Eu~~ $cov(X, u) = 0$ is not plausible in the ~~population~~ causal model ~~- it~~ because of omitted variables. X is not plausibly uncorrelated with unmodelled determinants of Y collected in u. For example, ~~a person~~ the quality of a person's diet is likely a ~~negative~~ positive determinant of Y, and is likely to be ~~positively~~ negatively correlated with X.

Consider the long causal model
$Y = \beta_0 + \beta_1 X + \beta_2 X' + v$,
where X' is the quality of a person's diet.
$\delta_1 = cov(Y, X)/var(X)$
$= cov(\beta_0 + \beta_1 X + \beta_2 X' + v, X)/var(X)$
$= \beta_1 + \beta_2\, cov(X', X)/var(X)$
$= \beta_1 + \beta_2 \pi_1$
where π_1 is the coefficient on X in the population regression of X' on X.

~~This is the om~~ The omitted variable bias is $\beta_2 \pi_1$. In general, X is likely to be positively correlated with ~~positive~~ negative determinants of Y and negatively correlated with ~~negative~~ positive determinants of Y so the omitted variable bias will be ~~positive~~ negative. ~~$\delta_1 > \beta_1$~~. $\delta_1 < \beta_1$.

OLS estimator $\hat{\delta}_1$ ~~for~~ is consistent for δ_1 and will ~~likely~~ ~~over~~ likely ~~overestimate~~ underestimate β_1.

b The OLS ~~estimator of the coeffici~~ coefficient on X in (B) is more plausibly consistent for β_1 because it eliminates one potentially large source of omitted variable bias.

Supposing that smoking habits W are the only relevant omitted variable in the determination of Y, the ~~causal mod~~ long causal model is
$Y = \beta_0 + \beta_1 X + \beta_2 W + v$,
where ~~Eu~~ $Ev = cov(X, v) = cov(W, v) = 0$ by supposition.
Then the causal model satisfies orthogonality and its coefficients coincide with those of the corresponding long population regression, and are consistently estimated by the long OLS regression. ~~If~~ Estimators from the short OLS regression will be biased as explained earlier.

The coefficient on X in OLS regression (A) is consistent for δ_1. This estimator is equal to $\hat{cov}(Y, X)/\hat{var}(X)$. By the consistency of $\hat{cov}$ and $\hat{var}$ for their population counterparts, this estimator is consistent for the population regression coefficient $\delta_1 = cov(Y, X)/var(X)$.

c The value of the estimated coefficient on X differs ~~from regression (A) to regr~~ between (A) and (B) because one source of omitted variable bias is eliminated in (B).

~~Smoking habit~~ W is plausibly a negative determinant of Y, because smoking ~~is bad for health~~ causes various diseases that reduce life expectancy. W is plausibly positively correlated with blood pressure because smoking undermines ~~health~~ heart health.

Suppose for simplicity that W is the only ~~releva~~ relevant omitted variable in the short population regression. Then, the long causal model is
$Y = \beta_0 + \beta_1 X + \beta_2 W + v$,
where $Ev = cov(X, v) = cov(W, v) = 0$ by supposition.
This satisfies orthogonality, so its coefficients coincide with those of the population linear regression
$Y = \gamma_0 + \gamma_1 X + \gamma_2 W + f$,
and are consistently estimated by the OLS estimators.
The population regression coefficient on X in the short population regression of Y on X alone is
$\delta_1 = cov(Y, X)/var(X)$
$= \beta_1 + \beta_2 \pi_1$
where π_1 is the coefficient on X in the ~~aux~~ auxiliary regression of W on X. The OLS estimator in (A) is consistent for δ_1. The OLS estimator in (B) is consistent for β_1. So the former ~~under~~ has a lower value than the latter because of negative omitted variable bias. Omitted variable bias is ~~likely to be~~ negative because, as explained above, W is a negative determinant of Y ($\beta_2 < 0$) and positively correlated with X ($\pi_1 > 0$).

Standard error is larger in (B) than in (A) because the variance of the component of X uncorrelated with W is significantly lower than the variance of ~~the~~ X itself. So even though, by construction, residuals of (B) have lower variance, the former effect dominates.

For homoskedastic ~~errors~~ residuals, ~~in (A)~~,
$se(\hat{\beta}^A) = n^{-1/2}\sqrt{s^2_{\hat{u}^A}/\widehat{var}(X)}$
$se(\hat{\beta}^B) = n^{-1/2}\sqrt{s^2_{\hat{u}^B}/\widehat{var}(\tilde{X})}$

$s^2_{\hat{u}^A} < s^2_{\hat{u}^B}$ but also $\widehat{var}(X) >> \widehat{var}(\tilde{X})$, so
$se(\hat{\beta}^B) > se(\hat{\beta}^A)$

d Suppose that orthogonality holds in the long causal model corresponding to (B), then the OLS coefficients of (B) are consistent for the parameters of the long causal model above. $\hat{\beta}_1 = -0.031 \xrightarrow{p} \beta_1$, $\hat{\beta}_2 = -2.131 \xrightarrow{p} \beta_2$. From the above, the OLS coefficient on X in (A) is consistent for $\delta_1 = \beta_1 + \beta_2\pi_1$. $\hat{\delta}_1 = -0.052 \xrightarrow{p} \delta_1$.
$\pi_1 = (\delta_1 - \beta_1)/\beta_2$, $\hat{\pi}_1 = (\hat{\delta}_1 - \hat{\beta}_1)/\hat{\beta}_2$
$= -0.052 - -0.031 / -2.131 = 0.0098545$. ~~0.0098~~
This is a consistent estimate of the ~~po~~ coefficient on X in a population linear ~~regr~~ regression of W on X. On average, a person with blood pressure higher by 1 mmHg consumes 0.0098545 more packs of cigarettes a day.

e Proposal (i) provides the best prediction. By construction of the population regression of Y on X alone, $\gamma_0 + \gamma_X X$ is the best ~~linear predictor~~ (mean-squared prediction error-minimising) predictor of Y given X alone. This is consistently estimated by $\hat{\gamma}_0 + \hat{\gamma}_X X$.

~~In~~ In contrast, the ~~pop~~ function $\hat{\gamma}_0^{(B)} + \hat{\gamma}_X^{(B)} X + \hat{\gamma}_W^{(B)} \bar{W}$ is a linear function of X that ~~is~~ is not consistent for the ~~=~~ short population regression above, so it is not consistent for the best linear predictor of Y given X alone.

The model in (i) effectively uses X to predict W and ~~computes an~~ predicts Y using X and this prediction of W. The model in ~~i~~ (ii) is lossy because it ~~does not use~~ fails to account for the effect of variability in W on Y.

f An estimate of the causal effect of X on Y is irrelevant to the insurance company that (presumably) is interested only in prediction and not in intervention. So a precise model ~~that~~ that ~~is~~ consistently estimates the descriptive relationship between X and Y is sufficient for the insurance company's purposes.

5a R^2 decreases from [A] to [B] to [C] because of the inclusion of additional regressors. By construction, the ~~ss~~ OLS regression coefficients minimise the SSR. Because ~~the~~ ~~AB~~ a model ~~model [A] can be replicated with ad~~ with fewer regressors can be replicated in a model with more regressors, the SSR of the larger model is always weakly lower than that of the smaller model, and strictly so if the additional regressors are even marginally relevant. So ~~the~~ SSR is decreasing from [A] to [B] to [C], hence $R^2 = 1 - SSR/TSS$ is increasing in that order.

The relative magnitude of R^2 between [C] and [D] is ambiguous, and depends on the relative degree to which the regressors between large and small winners account for variation in earnpost. Plausibly, R^2 of [D] is higher because large prize winner's post winning earnings are more unpredictable, and more subject to differences in personality.

b $H_0: \beta_{pr} = 0$
$H_1: \beta_{pr} < 0$

$t = -0.051 - 0 / 0.024 = -2.1250$

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 appropriate critical value drawn from the $N(0,1)$ distribution at the level of significance α.

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

$p = P(t < t^{act}) = \Phi(-2.1250) = 0.0168 = 1.68\%$

Under the null, the probability of observing a t statistic as unfavourable to the null as that actually observed is 1.68%. Reject the null at all levels of significance $\alpha > 1.68\%$. Fail to reject the null otherwise.

c The difference between $\hat{\beta}_{pr}^{(C)}$ and $\hat{\beta}_{pr}^{(D)}$ is significant at the $\alpha = 5\%$ level iff, at the $\alpha = 5\%$ level, there is a significant difference between the ~~pa~~ ~~£~~ earnpost response of small prizewinners to the earnpost response of large prizewinners.

Using the given data, estimate the following model by OLS

$$earnpost = \beta_0 + \beta_{spr}\, prize + \beta_{lpr-spr}\, [prize > 100{,}000]\, prize + \beta_e\, educ + \dots + \beta_{ep}\, earnpre + u$$

Perform the following hypothesis test.

$H_0: \beta_{lpr-spr} = 0$
$H_1: \beta_{lpr-spr} \neq 0$

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

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 appropriate critical value drawn from the $N(0,1)$ distribution at the $\alpha = 5\%$ level of significance. For a two-sided test, $\alpha = 1.960$.

If the null is rejected, conclude that there is a significant difference in the earnpost response of small prizewinners and large prizewinners, ~~FA~~ and that there is a significant difference between the reported estimates.

d Consider regression [C]. The estimate of β_{pr} from [C] fails to provide a reliable estimate of MPE from lottery winnings, i.e. the causal effect of lottery winnings on ~~earned~~ subsequent earned income iff there is endogeneity in the regressors or the causal effect of interest is non linear.

There is intuitive and empirical reason to think that the relationship between lottery winnings and subsequent earned income is non linear. Intuitively, earned income ~~is~~ has non-linearities because not all jobs allow for flexible hours and pay hourly. Plausibly a sufficiently large winning will prompt some to leave their jobs entirely. Empirically the large difference in estimates between [C] and [D] supports the hypothesis of non-linearity.

Given that ~~the sample is random~~ lottery winnings are random, they are uncorrelated with the unobserved determinants of subsequent earnings, so ~~#~~ prize is plausibly exogenous. There is no reason to suspect measurement error, and there will be no measurement error if the study is conducted ~~as~~ competently. Simultaneity is also not plausible because prize is determined temporally prior to earnpost. So prize is plausibly exogenous.

e This study has limited external validity because ~~the~~ of significant ~~exo~~ heterogeneity between the treatment in this study and the policy of interest, ~~lottery participants are likely to have ch~~ and between the population of this study and the population of interest.

First, lottery participants will be unlike ~~welfare~~ potential welfare recipients because the latter are more likely to be low income, and the former are more likely to have gambling habits ~~and / or~~ . Potentially, lower income individuals ~~with no in gambling habits~~ may have more incentive to earn income because of decreasing marginal utility of income.

A welfare payment is also unlike a lottery payout in that the welfare payment will likely cease once an ~~person has~~ individual has reached some level of income (or ~~decrease~~ otherwise decrease with increasing earned income) whereas the lottery payment will not. This potentially reduces the incentive to earn income.

f educ is an endogenous control. It is a post-treatment characteristic ~~that~~, and is likely to be correlated with unmodelled determinants of ~~earnpost~~ earnings. For example, educ is likely to be correlated with cognitive ability which is a positive determinant of earnings. Endogeneity in educ ~~implies undermines the~~ causes the estimated coefficient on prize to no longer be consistent for the causal effect of interest.

6a It is possible to estimate β_1 consistently by ILS or 2SLS iff ~~*~~ Z is a valid instrument for X. This is iff Z is relevant, i.e. $\pi_1 \neq 0$, Z is exogenous, i.e. $cov(Z, u) = 0$ (which is given), and Z is excluded, i.e. Z is not itself a direct determinant of Y (which is evident from the given causal model (structural equation).

ILS

Structural equation $Y = \beta_0 + \beta_1 X + \beta_2 W + u$

First stage regression $X = \pi_0 + \pi_1 Z + \pi_2 W + v$

By substitution, reduced form regression

$Y = \beta_0 + \beta_1(\pi_0 + \pi_1 Z + \pi_2 W + v) + \beta_2 W + u$

$= (\beta_0 + \beta_1\pi_0) + \beta_1\pi_1 Z + (\beta_2 + \beta_1\pi_2) W + (u + \beta_1 v)$

where $cov(W, u + \beta_1 v)$ holds by linearity of covariance and given that $cov(W, u) = cov(W, v) = 0$, and $cov(Z, u + \beta_1 v) = 0$ by linearity of covariance ~~and~~ by construction of the first stage regression, and given that $cov(Z, u) = 0$. Then, the reduced form regression is a population linear regression of Y on Z and W.

OLS coefficient on Z in the regression of Y on Z and W is consistent for $\beta_1\pi_1$. OLS coefficient ~~of~~ on Z in the regression of X on Z and W is consistent for π_1. The former divided by the latter is consistent for β_1.

2SLS

Estimate the first stage regression by OLS.

Compute $\hat{X}^* = \hat{\pi}_0 + \hat{\pi}_1 Z + \hat{\pi}_2 W$.

Estimate the second stage regression of Y on X^* and W by OLS.

The coefficient on X^* in the second stage regression is consistent for β_1.

b From the above,

~~Reduce~~

Reduced form regression

$Y = \gamma_0 + \gamma_1 Z + \gamma_2 W + \varepsilon$,

where $\gamma_0 = \beta_0 + \beta_1\pi_0$, $\gamma_1 = \beta_1\pi_1$, $\gamma_2 = \beta_2 + \beta_1\pi_2$, $\varepsilon = u + \beta_1 v$.

$\gamma_1 = \beta_1\pi_1$

$\gamma_1 = \beta_1 \iff \pi_1 = 1$

$\hat{\gamma}_1 \xrightarrow{p} \beta_1 \iff \pi_1 = 1$

$\hat{\gamma}_1$ is consistent for β_1 iff $\pi_1 = 1$, i.e. the coefficient on Z in the population linear regression of X on Z is 1.

c Short first stage regression

$X = \pi_0' + \pi_1' Z + v'$

$X^{*'} = \pi_0' + \pi_1' Z$

By substitution, second stage regression

$Y = \beta_0 + \beta_1(X^{*'} + v') + \beta_2 W + u$

$= \beta_0 + \beta_1 X^{*'} + \beta_2 W + (u + \beta_1 v')$

Except by coincidence, ~~cov(W,~~ where $\pi_2 = 0$ in the long first stage regression, $cov(W, v') \neq 0$ because X is correlated with W independently of Z, so $v' = X - \pi_0 - \pi_1 Z$ is correlated with W. Then, by linearity of covariance, $cov(W, u + \beta_1 v') = \beta_1 cov(W, v') \neq 0$. So W is endogenous in the ~~reduced form~~ ~~reg~~ second stage regression, the second stage regression fails to coincide with the population regression of Y on $X^{*'}$ and W, the coefficient on $\hat{X}^{*'}$ in the associated OLS regression is not consistent for β_1.

d $cov(Y, v)$

$= cov(\beta_0 + \beta_1 X + \beta_2 W + u, v)$

$= cov(\beta_1 X, v)$

$= \beta_1 cov(X, v)$

$= \beta_1 cov(\pi_0 + \pi_1 Z + \pi_2 W + v, v)$

$= \beta_1 cov(v, v)$

$= \beta_1 var(v)$. $\Rightarrow$

$\beta_1 = cov(Y, v)/var(v)$

Given that ~~cov(~~ $Ev = cov(Z, v) = cov(W, v) = 0$, the first stage regression $X = \pi_0 + \pi_1 Z + \pi_2 W + v$ is a population regression of X on Z and W, its coefficients are ~~consiste~~ consistently estimated by OLS, so $\hat{v} = X - \hat{\pi}_0 - \hat{\pi}_1 Z - \hat{\pi}_2 W$ is consistent for v.

β_1 is the coefficient on v in a population regression of Y on v alone. So the coefficient on $\hat{v}$ in a ~~p~~ OLS regression of Y on $\hat{v}$ is consistent for β_1.

7a ~~$E_t X_{t+k}$~~

~~$= E[X_{t+k} \mid X_t, X_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots]$~~

~~$= E[\phi X_{t+k-1}$~~

x_{t+k}

$= \phi x_{t+k-1} + u_{t+k}$

$= \phi(\phi x_{t+k-2} + u_{t+k-1}) + u_{t+k}$

$\vdots$

$= \phi^k x_t + u_{t+k} + \phi u_{t+k-1} + \phi^2 u_{t+k-2} + \dots + \phi^{k-1} u_{t+1}$

$= \phi^k x_t + \sum_{i=0}^{k-1} \phi^i u_{t+k-i}$

$= \phi^k x_t + \sum_{i=0}^{k-1} \phi^i (\varepsilon_{t+k-i} + \gamma \varepsilon_{t+k-i-1})$

$E_t x_{t+k}$

$= E[x_{t+k} \mid x_t, x_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots]$

$= E[\phi^k x_t + \sum_{i=0}^{k-1} \phi^i (\varepsilon_{t+k-i} + \gamma \varepsilon_{t+k-i-1}) \mid \cdots]$

$= \phi^k x_t + \sum_{i=0}^{k-1} \phi^i E[\varepsilon_{t+k-i} + \gamma \varepsilon_{t+k-i-1} \mid \cdots]$

$= \phi^k x_t + \sum_{i=0}^{k-1} (\phi^i E[\varepsilon_{t+k-i} \mid \cdots] + \phi^i \gamma E[\varepsilon_{t+k-i-1} \mid \cdots])$

$= \phi^k x_t + \phi^{k-1} \gamma \varepsilon_t$

$= \phi^{k-1} (\phi x_t + \gamma \varepsilon_t)$

by substitution, linearity of conditional expectation, conditioning, given that $\{\varepsilon_t\}$ is iid ~~so~~ and that ε_t is independent of $\{x_s\}_{s \le t-1}$, so ~~$E[\varepsilon_s]$~~ $E_t \varepsilon_s = E\varepsilon_s = 0$ for all $s > t$

b Price Y_t of the stock at t is equal to the present value of expected dividend payouts from t onward. This present value is the discounted sum of expected dividend payouts with discount factor δ between periods. A dividend payout of x in period $t+1$ has present value δx in period t.

ci ~~Given that $\{(x_t, \varepsilon_t)\}$ is weakly stationary, each of $\{x_t\}$ and $\{\varepsilon_t\}$ is weakly stationary, and the two c $\{x_t\}$ and $\{\varepsilon_t\}$ are jointly weakly stationary, the linear combination $E_t x_{t+k} = \phi^k x_t + \phi^{k-1} \gamma \varepsilon_t$ is weakly stationary.~~

$Y_t = \sum_{k=0}^{\infty} \delta^k E_t x_{t+k}$

$= \sum_{k=0}^{\infty} \delta^k \phi^{k-1} (\phi x_t + \gamma \varepsilon_t)$

$= 1/\phi \sum_{k=0}^{\infty} \delta^k \phi^k (\phi x_t + \gamma \varepsilon_t)$

$= 1/\phi (\phi x_t + \gamma \varepsilon_t) 1/(1 - \delta\phi)$

$= \frac{1}{\phi(1-\delta\phi)} (\phi x_t + \gamma \varepsilon_t)$

Y_t is a linear combination of x_t and ε_t, $\{x_t\}$ and $\{\varepsilon_t\}$ are jointly weakly stationary, so $\{Y_t\}$ is weakly stationary.

ii $\{Y_t\}$ Granger causes $\{x_t\}$ iff lags of $\{Y_t\}$ are useful for forecasting $\{x_t\}$. Formally, this is iff $E(x_t - E[x_t \mid \mathcal{X}_{t-1}])^2 = E(x_t - E[x_t \mid \mathcal{X}_{t-1}, \mathcal{Y}_{t-1}])^2$, which is iff $E[x_t \mid \mathcal{X}_{t-1}] = E[x_t \mid \mathcal{X}_{t-1}, \mathcal{Y}_{t-1}]$.

di $\phi = 1$, $x_0 = 0$

$Y_t = 1/(1-\delta) (x_t + \gamma \varepsilon_t)$

$= \frac{1}{1-\delta} x_t + \frac{\gamma}{1-\delta} \varepsilon_t$

Y_t and X_t are cointegrated with cointegrating coefficient ~~$1/(1-\delta)$~~. $\theta = 1/(1-\delta)$

$\varepsilon_t = Y_t - \theta X_t$

$= \frac{\gamma}{1-\delta} \varepsilon_t$

is stationary given that ε_t is iid hence stationary.

ii $\{W_t\}$ Granger causes $\{\Delta X_t\}$ iff

$E[\Delta X_t \mid \Delta X_{t-1}, \Delta X_{t-2}, \dots] = E[\Delta X_t \mid \Delta X_{t-1}, \dots, W_{t-1}, \dots]$

Conjecture that the relevant ψ is $1-\delta$.

$W_t = X_t - \psi Y_t$

$= \cancel{X_t - \frac{1-\delta}{1-\delta} X_t} - \gamma \varepsilon_t$

$E[\Delta X_t \mid \Delta X_{t-1}, \dots]$

~~$= E[X_t - X_t$~~

$= E[X_t \mid \Delta X_{t-1}, \dots] - E[X_{t-1} \mid \Delta X_{t-1}, \dots]$

$= E[X_t \mid X_{t-1}, X_{t-2}, \dots] - E[X_{t-1} \mid X_{t-1}, \dots]$

$= E[X_t \mid \mathcal{X}_t] - X_{t-1}$

$= E[\phi X_{t-1} + u_t \mid \mathcal{X}_t] - X_{t-1}$

$= \cancel{\phi} E[u_t \mid \mathcal{X}_t] - (1-\phi) X_{t-1}$

Date