Quantitative Economics Paper 180523

1a Consider the sample statistic $\hat{\theta}$ and the population parameter θ. The bias of $\hat{\theta}$ as an estimator of θ is $E(\hat{\theta}-\theta)$. $\hat{\theta}$ is an unbiased estimator of θ iff its bias as such, $E(\hat{\theta}-\theta)=0$.

Sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is an unbiased estimator of population mean $\mu_X = Ex_i$ where $\{x_i\}_{i=1}^{n}$ are ~~ind~~ iid with common expectation μ_X. + Prove that this is so.

Some ~~pop~~ sample statistic $\hat{\theta}$, computed from a sample of observations, is an unbiased estimator of the population parameter iff, intuitively, ~~the value of $\hat{\theta}$ is equal to the true~~ in expectation, the value of $\hat{\theta}$ is equal to θ. It makes sense to speak of the expectation of $\hat{\theta}$ because $\hat{\theta}$ is computed from a random sample, ~~and~~ so it is subject to sampling variation.

b The efficiency of an estimator of some population parameter is inversely related to the variability of the estimator (due to sampling variation). The greater the variability of an estimator, the less efficient it is. The efficiency of an estimator is measured (more precisely, estimated) by its standard error.

Consider the following two estimators of the population mean $\mu_X = EX$, $\hat{\mu}_{1,X} = x_1$, $\hat{\mu}_{n,X} = \frac{1}{n}\sum_{i=1}^{n} x_i$, given the iid sample $\{x_i\}_{i=1}^{n}$ for $n>1$. The latter estimator is more efficient because ~~its~~ it has lower variance.

$$var(\hat{\mu}_{1,X}) = var(x_1) = \sigma_X^2,$$
$$var(\hat{\mu}_{n,X}) = var(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i) = \tfrac{1}{n^2} var(\textstyle\sum_{i=1}^{n} x_i) = \sigma_X^2/n$$

where the first series of equalities follows entirely by substitution, and the second series of equalities follows by substitution, ~~the comm~~ common result ~~that $var(ax) = a^2 var(x)$,~~ for var, and given that $x_1, \ldots, x_n$ are independent.

c $\hat{\theta}$ is a consistent estimator for θ iff, as the size of the sample from which $\hat{\theta}$ is constructed becomes large, $\hat{\theta}$ converges in probability to θ.

~~$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$~~ ~~is~~ Sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is a consistent estimator of the population mean $\mu_X = EX$, ~~this result is the law~~ given that $\{x_i\}_{i=1}^{n}$ is an iid random sample. This result is the law of large numbers.

An estimator is consistent iff, as the sample size becomes large, its bias and variance ~~approach zero~~ converge to zero.

2 Let $e = Y - E[Y|X]$

$$
\begin{aligned}
E[e|X] &= E[Y - E[Y|X]|X] \\
&= E[Y|X] - E[E[Y|X]|X] \\
&= E[Y|X] - E[Y|X] \\
&= 0
\end{aligned}
$$

where the first equality follows by substitution, the second follows by linearity of conditional expectation, and the third follows by ~~EOA~~ conditioning ($E[Y|X]$ is a function of X and is "known" within the conditional expectation).

So e is mean-independent of X.

$$
\begin{aligned}
Ee &= E(Y - E[Y|X]) \\
&= EY - E[E[Y|X]] \\
&= EY - EY \\
&= 0
\end{aligned}
$$

by substitution, linearity of expectation, law of iterated expectations.

~~$E(e)$~~

$$
\begin{aligned}
E(eE[Y|X]) &= E(YE[Y|X] - E[Y|X]E[Y|X]) \\
&= E(YE[Y|X]) - E(E[Y|X]E[Y|X]) \\
&= EY^2 - EY^2 \\
&= 0
\end{aligned}
$$

By substitution, linearity of expectation, law of iterated expectations.

Then, $\mathrm{cov}(e, E[Y|X]) = E(eE[Y|X]) - EeE(E[Y|X]) = 0$

$\mathrm{var}(Y) = \cancel{\mathrm{var}(Y)}\ \mathrm{var}(E[Y|X]) + \mathrm{var}(e)$

given that $\mathrm{cov}(e, E[Y|X]) = 0$.

The required result is established.

exactly correct.

3a spurious regression is the tendency to find statistically significant relationships ~~as~~ between entirely independent time series with order of integration 1. "Y and X are both I(1)"

A time series $\{Y_t\}$ has order of integration 1 iff Y_t ~~has a unit root and $\Delta Y_t = Y_t - Y_{t-1}$ does not have a unit root. This is iff~~ is non-stationary and $\Delta Y_t = Y_t - Y_{t-1}$ is stationary.

Time series with order of integration 1 consist of random variables with a stochastic trend. Such time series therefore tend to exhibit large swings of increase and decrease, which, with surprising regularity, can be matched to similar swings in unrelated time series with order of integration 1.

If, for example, some statistically significant relationship is found between the number of Oxford graduates in a year and the ~~Acurabe~~ size of the population of squirrels in Canada, such a relationship is likely to be spurious.

b If ~~two~~ Y_t and X_t are cointegrated, then any statistically significant relationship found between the time series $\{Y_t\}$ and $\{X_t\}$, even if the two ~~a~~ ~~have~~ time series have order of integration 1, is genuine and not spurious.

Y_t and X_t are cointegrated ~~iff~~ iff there exists some cointegrating coefficient θ such that ~~$\varepsilon_t = Y_t - \theta X_t$ is~~ equilibrium error $\varepsilon_t = Y_t - \theta X_t$ is stationary.

It is possible to test for cointegration by first estimating θ by OLS regression of Y_t on X_t (or using some hypothesised value of θ), then computing the estimated equilibrium error (or ~~using~~ the equilibrium error obtained using the hypothesised value of θ) and performing an (augmented) Dickey Fuller test for a unit root in the estimated equilibrium error. If the null of a unit root is rejected, conclude that the equilibrium error is stationary and that Y_t and X_t are in fact cointegrated with the estimated (or hypothesised) cointegrating coefficient, and that the statistically significant relationship between the two is ~~go~~ genuine rather than spurious.

Alternatively, ~~a regression of~~ the first stationary difference of each of Y_t and X_t ~~is~~ has order of integration 0, so a regression of the first stationary difference of Y_t on X_t (or the first stationary difference of X_t) is immune to the problem of spurious regression.

Date

a Let f and F denote respectively the probability density function and the cumulative distribution function of X_i. ~~resp~~

$$f(x) = \begin{cases} p & \text{for } x = 1 \\ 1-p & \text{for } x = 0 \\ 0 & \text{for all other } x \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1-p & \text{for } 0 \leq x < 1 \\ 1 & \text{for } 1 \leq x \end{cases}$$

$f(x) = P(X_i = x)$ for all x, $F(x) = P(X_i \leq x)$ for all x,
$F(x) = \int_{-\infty}^{x} f(y)\,dy$

b
$$\begin{aligned} EX_i &= \int_{-\infty}^{\infty} x f(x)\,dx \\ &= 1 \times p + 0 \times (1-p) \\ &= p \end{aligned}$$

$$\begin{aligned} var(X_i) &= E(X_i - EX_i)^2 \\ &= E(X_i - p)^2 \\ &= E(X_i^2 - 2pX_i + p^2) \\ &= p^2 + E(X_i^2) - 2pEX_i \\ &= p^2 + \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2p^2 \\ &= -p^2 + (1^2 \times p + 0^2 \times 1-p) \\ &= p - p^2 \\ &= p(1-p) \end{aligned}$$

by definition of var, substitution of earlier result, expansion, linearity of expectation, definition of expectation.

c $\hat{p} = n^{-1} \sum_{i=1}^{n} X_i$

$$\begin{aligned} var(\hat{p}) &= var\left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} X_i\right) \\ &= \tfrac{1}{n^2} var\left(\textstyle\sum_{i=1}^{n} X_i\right) \\ &= p(1-p)/n \end{aligned}$$

by substitution, common results for var, given that $X_1, \ldots, X_n$ are iid and each has ~~Ber~~ Bernoulli (p) distribution.

$$\begin{aligned} sd(\hat{p}) &= \sqrt{var(\hat{p})} \\ &= \sqrt{p(1-p)/n} \end{aligned}$$

$$se(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$$

Given that $\hat{p}$ is a consistent estimator of p.

$$\begin{aligned} E\hat{p} &= E\left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} X_i\right) \\ &= \tfrac{1}{n} \textstyle\sum_{i=1}^{n} EX_i \\ &= \tfrac{1}{n} np \\ &= p \end{aligned}$$

As $n \to \infty$, $var(\hat{p}) \to 0$.

$\hat{p}$ is an unbiased estimator of p and $var(\hat{p})$ converges to 0 as n becomes large, so $\hat{p}$ is a consistent estimator of p.

This proof is important.

d $\{X_i\}_{i=1}^{n}$ is iid. $EX_i = p$, $var(X_i) = p(1-p)$ for all i. $\hat{p} = \bar{X} = \tfrac{1}{n}\sum_{i=1}^{n} X_i$.

By CLT, the standardised sample mean $(\bar{X} - EX_i)/\sqrt{var(X_i)/n} = (\hat{p} - p)/\sqrt{p(1-p)/n}$ converges in distribution to the standard normal $N(0,1)$, as n becomes large.

$se(\hat{p})$ is obtained by substituting the population parameter p in $sd(\hat{p}) = \sqrt{p(1-p)/n}$ for its consistent estimator $\hat{p}$. ~~so it is~~ So it is approximately true that $(\hat{p} - p)/se(\hat{p})$ converges in distribution to the standard normal.

e $se(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0.3 \times 0.7/100} = 0.045826$

By CLT, given that the sample is an iid random sample, and that the sample is large, the standardised sample mean (and its estimate computed using ~~sd~~ $se(\hat{p})$ rather than $sd(\hat{p})$) converges in distribution to the standard normal.

Explain that this is a t-statistic.

The required confidence interval is

$$\begin{aligned} C &= [\hat{p} - 1.960\, se(\hat{p}),\ \hat{p} + 1.960\, se(\hat{p})] \\ &= [0.21018,\ 0.38982] \end{aligned}$$

With 95% probability, the random interval C contains the true value of the population parameter p. At the 5% level of significance, ~~Both~~ reject any null outside this interval.

Date

5ai Consider the following OLS regression.

$$\ln x_h = \hat{\beta}_0 + \hat{\beta}_1 \ln x_w + \hat{u}$$

By construction, $\hat{\beta}_1 = \widehat{cov}(\ln x_h, \ln x_w)/\widehat{var}(\ln x_w)$
$= \widehat{cov}(\ln x_h, \ln x_w)/\widehat{sd}^2(\ln x_w)$

In 1960, $\hat{\beta}_1 = 10/5^2 = 0.4$

By similar calculations,

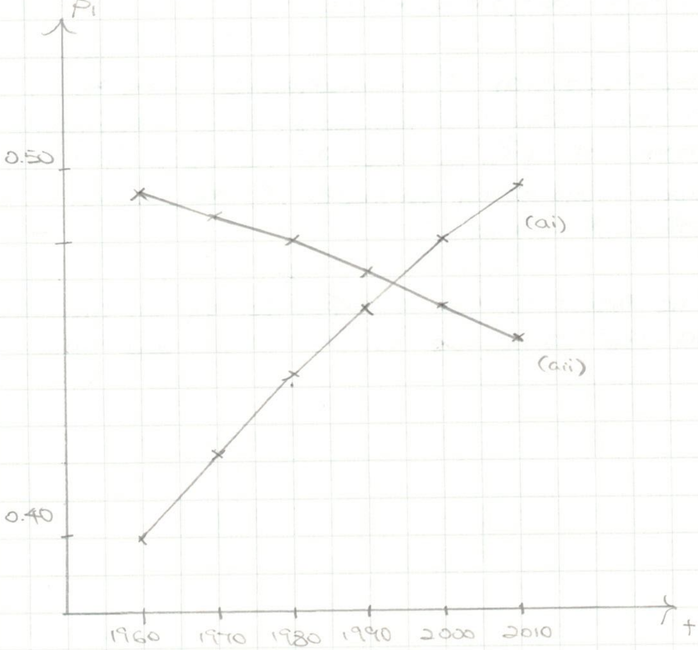
Year	$\hat{\beta}_1$
1960	0.40000
1970	0.42291
1980	0.44379
1990	0.46280
2000	0.48011
2010	0.49587

ii Consider the following OLS regression

$$\ln x_w = \hat{\beta}_0 + \hat{\beta}_1 \ln x_h + \hat{u}$$

By construction, $\hat{\beta}_1 = \widehat{cov}(\ln x_w, \ln x_h)/\widehat{var}(\ln x_h)$
$= \widehat{cov}(\ln x_w, \ln x_h)/\widehat{sd}^2(\ln x_h)$

Year	$\hat{\beta}_1$
1960	0.49383
1970	0.48753
1980	0.48000
1990	0.47166
2000	0.46281
2010	0.45369



c Consider the following population linear regressions.

$$y = \alpha + \beta_m \ln x_m + \beta_f \ln x_f + \gamma X + e$$
$$\ln x_m = \delta_0 + \delta_f \ln x_f + \delta_x X + \widetilde{\ln x_m}$$
$$\ln x_f = \pi_0 + \pi_m \ln x_m + \pi_x X + \widetilde{\ln x_f}$$

~~β_m~~ β_m is the OLS estimand of $\hat{\beta}_m$ and β_f is the OLS estimand of $\hat{\beta}_f$. By construction of the above regressions, $\beta_m = cov(y, \widetilde{\ln x_m})/var(\widetilde{\ln x_m})$ and $\beta_f = cov(y, \widetilde{\ln x_f})/var(\widetilde{\ln x_f})$. This is an application of the Frisch-Waugh-Lovell theorem.

Whether each estimand coincides with the associated effect depends on whether each regressor (of $\ln x_m$ and $\ln x_f$) is endogenous. There is no reason to suspect measurement error in either of these regressors. Simultaneity is implausible because x_m and x_f are realised temporally before y, so it cannot be that y is a determinant of either $\ln x_m$ or $\ln x_f$. The only ~~pot~~ plausible potential source of endogeneity is omitted variable bias. Then, $\ln x_m$ (and/or $\ln x_f$) is endogenous ~~=~~ iff it is correlated with unmodelled determinants of y. It is not plausible that all determinants of y potentially correlated with $\ln x_m$ (and/or $\ln x_f$) are captured (or at least adequately proxied for) in X. For example, it is not clear that a measure of or ~~proxy~~ appropriate proxy for social capital is available. So $\ln x_m$ and $\ln x_f$ are not plausibly exogenous. Then, the ~~e~~ OLS estimands (population regression parameters) fail to coincide with the relevant causal effects.

d ~~Conduct~~ Perform the following F test.

$H_0: \beta_m = \beta_f = 0$
$H_1: \beta_m \neq 0$ or $\beta_f \neq 0$

~~Es Unrestricted OLS~~

Estimate the unrestricted linear regression model by OLS.

$$y = \hat{\alpha} + \hat{\beta}_m \ln x_m + \hat{\beta}_f \ln x_f + \hat{\gamma} X + \hat{e}$$

Estimate the restricted linear regression model

$$y = \hat{\alpha}' + \hat{\gamma}' X + \hat{e}'$$

Compute the respective sum of squared errors.

$SSR = \sum_{i=1}^{n} \hat{e}_i$, $SSR' = \sum_{i=1}^{n} \hat{e}_i'$

Compute the F-statistic.

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

where ~~k=2 is the~~ $k*$ is the number of regressors (and so depends on the number of dimensions of x) and $q=2$ is the number of restrictions.

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

If the null is rejected, conclude that parental log wealth predicts children's log wealth.

- Consider the auxiliary population linear ~~regress~~ regression.

$$\ln x_f = \delta_0 + \delta_1 \ln x_m + \delta_2 x + v.$$

Substitute into the population linear regression of y on $\ln x_m$, $\ln x_f$, and x.

$$y = \alpha + \beta_m \ln x_m + \beta_f(\delta_0 + \delta_1 \ln x_m + \delta_2 x + v) + \gamma' x + u$$
$$= (\alpha + \beta_f \delta_0) + (\beta_m + \beta_f \delta_1)\ln x_m + (\gamma + \beta_f \delta_2)x + (\beta_f v + u)$$

~~Suppose for simplicity that~~

By construction of the auxiliary regression, v is uncorrelated with $\ln x_m$ and x. Then, by linearity of covariance, $\beta_f v + u$ is uncorrelated with $\ln x_m$ and x. By ~~a similar~~ construction of the auxiliary regression and by linearity of ~~ex~~ expectation, $E(\beta_f v + u) = 0$, so the above is a population linear regression of y on $\ln x_m$ and x.

$\overset{\circ}{\beta}_m$ is consistent for $\beta_m + \beta_f \delta_1$. $\hat{\beta}_m$ is consistent for β_m. (By the consistency of OLS estimators). The difference between $\overset{\circ}{\beta}_m$ and $\hat{\beta}_m$ is consistent for the omitted variable bias, which is the product of the coefficient on $\ln x_f$ in the ~~pa~~ "long" regression and the coefficient on $\ln x_m$ in the auxiliary regression.

From (a), the latter coefficient is likely to be positive. Intuitively, the former coefficient is also likely to be positive. So $\overset{\circ}{\beta}_m$ likely "overestimates" $\hat{\beta}_m$.

b Suppose that x_h is observed with error such that $\ln x_h$ has random error ε. Suppose favourably that this error is ~~random noise~~ ~~GA~~ noise and is uncorrelated with $\ln x_m$ or $\ln x_h$.

Consider the population regression of ~~$\ln x_h$~~ $\ln x_u$ on $\ln x_h$.

$$\ln x_u = \beta_0 + \beta_1 \ln x_h + u$$
$$= \beta_0 + \beta_1(\widetilde{\ln x_h} - \varepsilon) + u$$
$$= \beta_0 + \beta_1 \widetilde{\ln x_h} + (-\beta_1 \varepsilon + u).$$

~~The OLS estimator computed from ... of the coefficient on $\widetilde{\ln x_h}$ in an OLS regression of $\ln x_u$ on $\widetilde{\ln x_h}$ is ... $\hat{\beta}_1 = \widehat{cov}(\ln x_u, \widetilde{\ln x_h})/$~~

~~$\tilde{\beta}_1 = \widehat{cov}(\ln x_u, \widetilde{\ln x_h}) / var(\widetilde{\ln x_h})$~~
~~$= \widehat{cov}($~~

The coefficient on $\widetilde{\ln x_h}$ in a population linear regression of $\ln x_u$ on $\widetilde{\ln x_h}$ is

$$\tilde{\beta}_1 = cov(\ln x_u, \widetilde{\ln x_h}) / var(\widetilde{\ln x_h})$$
$$= cov(\beta_0 + \beta_1 \widetilde{\ln x_h} - \beta_1 \varepsilon + u, \widetilde{\ln x_h}) / var(\widetilde{\ln x_h})$$
$$= \beta_1 var(\widetilde{\ln x_h}) - \beta_1 var(\varepsilon) / var(\widetilde{\ln x_h})$$
$$= \beta_1 [var(\widetilde{\ln x_h}) - var(\varepsilon) / var(\widetilde{\ln x_h})]$$
$$= \beta_1 [var(\ln x_h) / var(\ln x_h) + var(\varepsilon)]$$

The OLS ~~esti~~ coefficient on $\widetilde{\ln x_h}$ in the OLS regression of $\ln x_u$ on $\widetilde{\ln x_h}$ is consistent for $\tilde{\beta}_1$.

~~Consider the population linear regression of $\widetilde{\ln x_h}$ on $\ln x_u$.~~

$$\widetilde{\ln x_h} = \beta_0 + \beta_1 \ln x_u + u + \varepsilon$$

By supposition ε is uncorrelated with $\ln x_u$, so the above is a population linear regression model of $\widetilde{\ln x_h}$ on $\ln x_u$. OLS regression of $\widetilde{\ln x_h}$ on $\ln x_u$ consistently estimates β_1, the causal effect of interest.

The result in (a) is explained by an increase in $var(\varepsilon)$ relative to $var(\ln x_h)$ over time.

6a On average, a woman in the population from which the given sample is drawn had a log wage lower than a man in this population by 0.17. Equivalently, on average, a woman in this population had a wage ~~0.8436~~ $e^{-0.17} = 0.84366$ times that of a man in this population.

important

The required confidence interval is

$$C = [-0.17 - 2.576(0.03), -0.17 + 2.576(0.03)]$$
$$= [-0.24728, -0.09272]$$

The random interval C contains the true value of the corresponding population coefficient with 99% probability. At the 1% level of significance, reject any null hypothesis of a population coefficient outside this interval.

b Age is included as a proxy for work experience, which it make sense to control for given that work experience is a potential source of omitted variable bias.

Education is included as a ~~proxy~~ control for the same reason.

If these controls are not included, then if gender is correlated with either work experience or ~~edu~~ degree status, then gender is correlated with the unobserved determinants of ~~w~~ (log) wage, so it is endogenous, and OLS ~~regressi~~ estimation of the coefficient on gender in the "short" regression is not consistent for the causal effect of gender on (log) wage.

apparently sufficient

$H_0: \beta_1 = -0.17$
$H_1: \beta_1 \neq -0.17$

* where β_1 is the coefficient on woman in the population linear regression corresponding to (2).

t-statistic
$t = \frac{-0.13 - -0.17}{0.03} = 1.3333.$

~~Reject the null if $|t| \geq c_\alpha$, where~~

Under the null, by CLT, given a sufficiently large iid random sample, ~~$\hat{\beta}$~~ t is approximately distributed according to the standard normal distribution.

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.

$P(|t| > c_\alpha) = \alpha \Leftrightarrow$ ~~c_α~~ $\alpha = 2\Phi(-c_\alpha) \Leftrightarrow$
$c_\alpha = 2.576.$ 1.645

Fail to reject the null. ~~that $\beta_1 =$~~ Fail to conclude that the coefficient on woman in the population regression corresponding to (2) is -0.17.

c Agree. Incorrect "compositional bias" out of syllabus

Omission of occupation dummies from regression (2) results in omitted variable bias, ~~but~~ and endogeneity of woman, then, the OLS estimator of the coefficient on woman in (2) is not consistent for the causal effect of interest, which is the direct causal effect of ~~gende~~ woman on (log) wage.

Omission of occupation dummies results in omitted variable bias because occupation is a potentially plausible direct determinant of (log) wage, and it is plausible that woman is correlated with such occupation dummies. For example, it could be the ~~ca~~ case that women disproportionately are employed in service sector roles.

Consider the simple case of a single occupation dummy W. ~~Suppose that~~ The relevant causal model is

$$\ln(wage) = \beta_0 + \beta_1 woman + \beta_2 age + \beta_3 education + \beta_4 W + u$$

~~The consider~~ The auxiliary population regression of W on woman ~~#~~ (and controls) is

$$W = \pi_0 + \pi_1 woman + \pi_2 age + \pi_3 education + v$$

By substitution,

$$\ln(wage) = (\beta_0 + \beta_4\pi_0) + (\beta_1 + \beta_4\pi_1) woman + (\beta_2 + \beta_4\pi_2) age + (\beta_3 + \beta_4\pi_3) education + (\beta_4 v + u)$$

By construction of the auxiliary regression, v is mean-zero and uncorrelated with woman, age, and education. Then, by linearity of expectation and covariance, $\beta_4 v + u$ is mean-zero and uncorrelated with woman, age, and education (supposing that orthogonality holds in the "long" causal model).

So the above is a population regression model of ln(wage) on woman, age, and education

8/10

The coefficients of the corresponding OLS regression are consistent for these. In particular, the OLS coefficient on woman is consistent for $\beta_1 + \beta_4 \pi_1$ rather than the causal effect of interest β_1.

Given that woman is correlated with the occupation dummy, $\pi_1 \neq 0$ and given that ~~the~~ occupation ~~dummies are~~ is a determinant of wage, $\beta_4 \neq 0$, so $\beta_1 + \beta_4 \pi_1 \neq \beta_1$.

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

$t = {-0.01 - 0}/{0.04} = -0.25000$

Under the null, given a sufficiently large iid random sample, by CLT, t is approximately distributed according to the standard normal distribution.

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

$p = P(|N(0,1)| > |t|)$
$= 2\Phi(-|t|)$
$= 2\Phi(-0.25000)$
$= 2(0.4013)$
$= 0.8026$

Under the null, the probability of observing a t statistic at least as unfavourable to the null as that actually observed is large, 0.8026. Fail to reject the null at any reasonable level of significance ($\alpha < 0.8026$).

impt

The ~~coeffi~~ OLS coefficients in (3) suggest that controlling for height, there is no difference between the (log) wage of men and women in the sample. ~~Possi~~ One explanation of height's accounting for the average difference in wage between men and women is that height is correlated with physical strength, confidence, and / or other such positive determinants of (log) wage.

2a $\cancel{y_t}$ $y_{-1} = y_0 = 0$

$y_t = \phi y_{t-2} + u_t$

$\{u_t\}$ is iid, $Eu_t = 0$, $var(u_t) = \sigma^2$

$|\phi| = 1$

Suppose $\phi = 1$

$y_1 = y_{-1} + u_1$

$Ey_1 = E(y_{-1} + u_1)$

$= Ey_{-1} + Eu_1$

$= 0 + 0$

$var(y_1) = var(y_{-1} + u_1)$

$= var(y_{-1}) + var(u_1)$

$= \cancel{2t}\ \sigma^2$

where the second inequality follows given that u_t is independent of y_s for all $s<t$.

$y_2 = y_0 + u_2$

$Ey_2 = E(y_0 + u_2)$

$= Ey_0 + Eu_2$

$= 0$

$var(y_2) = var(y_0 + u_2)$

$= var(y_0) + var(u_2)$

$= \sigma_2$

$y_3 = y_1 + u_3$

$Ey_3 = E(y_1 + u_3)$

$= Ey_1 + Eu_3$

$= 0$

$var(y_3) = var(y_1 + u_3)$

$= var(y_1) + var(u_3)$

$= \cancel{\sigma^2 + \sigma^2}\ 2\sigma^2$

Suppose $\phi = -1$,

$Ey_1 = -Ey_{-1} + Eu_1 = 0$

$var(y_1) = var(y_1) + var(u_1) = \cancel{2}\ \sigma^2$

$Ey_2 = -Ey_0 + Eu_2 = 0$

$var(y_2) = var(y_0) + var(u_2) = \sigma^2$

$Ey_3 = -Ey_1 + Eu_3 = 0$

$var(y_3) = var(y_1) + var(u_3) = \sigma^2$

~~Each~~ The series $\{y_t\}$ can be understood as composed of two alternating stochastic trends. The series $\{y_t\}$ is non-stationary because each of these stochastic trends is non-stationary. In particular, the variance of each trend increases ~~ever~~ in each (two) period (interval).

bi $|\phi| < 1$

$y_{-1} \perp\!\!\!\perp y_0$

$Ey_{-1} = Ey_0 = 0$

$var(y_{-1}) = var(y_0) = \sigma^2/1-\phi^2$

Supposing that "$\{y_t\}$" refers to $\{y_t\}_{t=-1}^T$ ~~?~~ or to $\{y_t\}_{t=0}^T$, given that $\{y_t\}$ is stationary, the mean and variance of y_t are simply the mean and variance of y_{-1} or y_0, namely 0 and $\sigma^2/1-\phi^2$ respectively. Correct

Suppose instead that $\{y_t\}$ is instead $\{y_t\}_{t=1}^T$.

$\cancel{y_t}$

$Ey_t = E\phi y_{t-2} + u_t)$

$= \phi E(y_{t-2}) + E(u_t)$

$= \phi Ey_{t-2}$

$= \phi^2 Ey_{t-4}$

$\vdots$

$= \cancel{Ey_0 \text{ or } Ey_{-1}}$ $\phi^{t/2} Ey_0$ for even t

$\phi^{t+1/2} Ey_{-1}$ for odd t

$= 0$

by substitution, linearity of expectation, given that $Eu_t = 0$ and by recursive substitution.

$var(y_t) = var(\cancel{y_{t-2}}\ \phi y_{t-2} + u_t)$

$= \cancel{var}\ \phi^2 var(y_{t-2}) + \sigma^2$

$= \phi^2(\phi^2 var(y_{t-4}) + \sigma^2) + \sigma^2$

$= \cancel{\phi^t var(y_0) + \sigma^2/1-\phi^2}$

$$\begin{cases} \phi^t var(y_0) + (1-\phi^t)\sigma^2/1-\phi^2 & \text{for even } t \\ \phi^{t+1} var(y_{-1}) + (1-\phi^{t+1})\sigma^2/1-\phi^2 & \text{for odd } t \end{cases}$$

$= \sigma^2/1-\phi^2$

ii $\cancel{cov(y_t, y_{t-1})}$

$\cancel{= cov(\phi y_{t-2} + u_t, \phi y_{t-3} + u_{t-1})}$

$\cancel{= cov(\phi^2 y_{t-4} + \phi u_{t-2} + \phi u_t, \phi^2 y_{t-3} + \phi u_{t-3} + u_{t-1})}$

$\cancel{= cov(\phi^{t/2} y_0 + \Sigma_{i=0}^{t/2-1} \phi^i u_{t-2i}, \phi^{t+1/2} y_{-1} + \Sigma_{i=0}^{t+1/2-1} \phi^i u_{t-2i-1}}$

~~consider odd even t~~

$\cancel{cov}\ y_t = \phi y_{t-2} + u_t$

$= \phi^2 y_{t-4} + \phi u_{t-2} + u_t$

$\vdots$

~~consider odd t~~

$\cancel{y_t =}$

~~Each odd~~ For all odd t, y_t is a function of y_{-1} and all us for odd $s \geq 0$. For all even t, y_t is a function of y_0 and us for even $s > 0$

Given that (1) y_{-1} and y_0 are independent, (2) $\{u_s\}$ is iid, and (3), ~~the~~ each u_s ~~is~~ for $s>0$ is independent of y_0, y_{-1}, every odd y_t is independent of every even y_t. ~~so cov~~

So ~~cov(t, t-1)~~ $cov(y_t, y_{t-1}) = 0$

iii For odd h, ~~*~~ either t is odd and $t-h$ is even, or t is even and $t-h$ is odd. Then, by the result above, y_t and y_{t-h} are independent, so $cov(y_t, y_{t-h}) = 0$ and $\rho_h = 0$.

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

$= cov(\phi y_{t-2} + u_t, y_{t-h})$

$= cov(\phi^2 y_{t-4} + \phi u_{t-2} + u_t, y_{t-h})$

$\vdots$

$= cov(\phi^{h/2} y_{t-h} + \phi^{h/2-1} u_{t-h+2} + \cdots, y_{t-h})$

$= \text{cov}(\phi^{h/2} y_{t-h}, y_{t-h})$

(given that $u_s \perp\!\!\!\perp y_t$ for all $s > t$, in particular that $u_{t-h+2}, u_{t-h+4}, \dots \perp\!\!\!\perp y_{t-h}$)

$= \phi^{h/2} \text{var}(y_{t-h})$

$\rho_h = \phi^{h/2} \text{var}(y_{t-h}) / \text{var}(y_t)$

$= \phi^{h/2} \text{var}(y_{t-h}) / \text{var}(y_{t-h})$

(given stationarity

$= \phi^{h/2}$

c $E[y_{t+1} | y_t]$

$= E[\phi y_{t-1} + u_{t+1} | y_t]$

$= E[\phi y_{t-1} | y_t] + E[u_{t+1} | y_t]$

$= \phi y_{t-1} + E u_{t+1}$

$= \phi y_{t-1}$

by substitution, linearity of conditional expectation, conditioning, given that $u_{t+1} \perp\!\!\!\perp y_t, y_{t-1}, \dots,$ and given that $E u_{t+1} = 0$.

$E[y_{t+2} | y_t]$

~~$= E[y_t | y_t]$~~

$= E[\phi y_t + u_{t+2} | y_t]$

$= \phi E[y_t | y_t] + E[u_{t+2} | y_t]$

$= \phi y_t$

by exactly the same steps.

$E[y_{t+3} | y_t]$

$= E[\phi y_{t+1} + u_{t+3} | y_t]$

$= E[\phi^2 y_{t-1} + \phi u_{t+1} + u_{t+3} | y_t]$

$= \phi^2 E[y_{t-1} | y_t] + \phi E[u_{t+1} | y_t] + E[u_{t+3} | y_t]$

$= \phi^2 y_{t-1} + \phi E u_{t+1} + E u_{t+3}$

$= \phi^2 y_{t-1}$

by substitution, linearity of ~~expectation expectation~~ conditional expectation, conditioning, given that $u_{t+1}, u_{t+3} \perp\!\!\!\perp y_t$, and that $E u_{t+1} = E u_{t+3} = 0$

$E[y_{t+4} | y_t] = \phi^2 y_t$

by exactly the same steps.

d $E[y_{t+1} | y_t] = E y_{t+1} = 0$

where the first equality follows by independence of y_{t+1} and y_t (which follows from y_{t+1} being a function of odd (even) ~~u~~ us and y_{-1} (y_0) and y_t being a function of even (odd) us and y_0 (y_{-1})), and the second equality follows by substitution of the earlier result.