Quantitative Economics Paper 270529

The required probability is 0.7888.

1a $E\bar{Y} =_1 E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$

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

$=_3 \frac{1}{n}\sum_{i=1}^{n} [p(1)+(1-p)(0)]$

$=_4 \frac{1}{n}\sum_{i=1}^{n} p$

$=_5 p$

where $=_1$ follows by substitution, $=_2$ by linearity of expectation, $=_3$ by definition of expectation, $=_4$ and $=_5$ by basic arithmetic

$var(\bar{Y}) =_1 var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$

$=_2 \frac{1}{n^2} var\left(\sum_{i=1}^{n} Y_i\right)$

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

$=_4 \frac{1}{n^2}\sum_{i=1}^{n} E(Y_i - EY_i)^2$

$=_5 \frac{1}{n^2}\sum_{i=1}^{n} E(Y_i^2 - 2pY_i + p^2)$

$=_6 \frac{1}{n^2}\sum_{i=1}^{n} (EY_i^2 - 2pEY_i + p^2)$

$=_7 \frac{1}{n^2}\sum_{i=1}^{n} (p - 2p + p^2)$

$=_8 p(1-p)/n$

where $=_1$ follows by substitution, $=_2$ by the common result, $=_3$ by the common result, given that $Y_1, \dots, Y_n$ are iid hence $cov(Y_i, Y_j) = 0$ for $i \neq j$, $=_4$ by definition of variance, $=_5$ by expansion and substitution, $=_6$ by linearity of expectation, $=_7$ by substitution and definition of expectation, and $=_8$ by basic arithmetic.

b By CLT, given that $Y_1, \dots, Y_n$ are iid, $(\bar{Y} - E\bar{Y})/var(\bar{Y})$ converges in distribution to the standard normal $Z \sim N(0,1)$ as n becomes large.

c Let m be the number of defective lightbulbs after one year.

~~$P\{15 \le m \le 25\} = P\{0.15 \le \bar{Y} \le 0.25\} = P\{ \frac{0.15 - E\bar{Y}}{\sqrt{var(\bar{Y})}} \le \bar{Y} \le \frac{0.25 - E\bar{Y}}{\sqrt{var(\bar{Y})}} \}$~~

$E\bar{Y} = p = 0.2$, $var(\bar{Y}) = p(1-p)/n = 0.2(0.8)/100 = 0.0016$

$P\{15 \le m \le 25\} = P\{15 \le n\bar{Y} \le 25\}$

$= P\{0.15 \le \bar{Y} \le 0.25\}$

$= P\left\{ \frac{0.15 - E\bar{Y}}{\sqrt{var(\bar{Y})}} \le \frac{\bar{Y} - E\bar{Y}}{\sqrt{var(\bar{Y})}} \le \frac{0.25 - E\bar{Y}}{\sqrt{var(\bar{Y})}} \right\}$

$= P\left\{-1.25 \le \frac{\bar{Y} - E\bar{Y}}{\sqrt{var(\bar{Y})}} \le 1.25\right\}$

$\approx P\{-1.25 \le Z \le 1.25\}$

$= P\{Z \le 1.25\} - P\{Z \le -1.25\}$

$= 0.8944 - 0.1056$

$= 0.7888$

2a Supposing that $\{Y_t\}$ is stationarity, each Y_t has common mean $\mu = EY_t$ and common variance ~~≠~~ $\sigma^2 = var(Y_t)$. Then, the ~~minimal~~ MSFE minimisi forecast of Y_{t+1} given only Y_t is simply Y_t.

$$\cancel{E(Y_{t+1}-Y_t)^2 = E(Y_{t+1}^2 - 2Y_tY_{t+1} + Y_t^2)}$$
$$\cancel{= E(Y_{t+1}^2) - 2Y_tE(Y_{t+1}) + Y_t^2}$$

$$E(Y_{t+1} - X_t)^2 = E(Y_{t+1}^2 - 2X_tY_{t+1} + X_t^2)$$
$$= EY_{t+1}^2 - 2X_tEY_{t+1} + X_t^2$$

FOC: $-2EY_{t+1} + 2X_t = 0 \Rightarrow X_t = EY_{t+1} = \mu$
SOC: $2 > 0$

The MSFE-minimising forecast of Y_{t+1} is μ.

Supposing $\{Y_t\}$ is stationary, Y_t is a consistent and ~~BASE~~ MSE minimising estimator for μ.

Then, given only Y_t, the MSFE minimising foreca of Y_{t+1} is Y_t.

b The MSFE-minimising forecast of Y_{t+1} given only ~~Y_t~~ Y_t is equal to $Y_t + c$, where c is some consta (where c is non-zero) if ~~EY_t~~ Y_t has a ~~determ~~ deterministic trend ~~and increases~~ or a ~~stochast~~ stochastic trend with mean c.

One such variable is the logarithm of price levels, which increases by an amount proportionate to the inflation rate in each period.

Formally, the condition is $E[Y_{t+1}|Y_t] = Y_t + c$

Show $E[Y_{t+1}|\cancel{Y_t}\ y_t]$ minimises ~~$E[Y_{t+1}|Y_t]$~~
$E[Y_{t+1} - Y_{t+1}|t]^2$

3ii $\hat{u}_i = w_i - \hat\beta_0 + \hat\beta_x x_i + \hat\beta_c c_i + \hat\beta_T T_i$

where, by construction of the OLS regression, $\widehat{cov}(\hat u_i, x_i) = \widehat{cov}(\hat u_i, c_i) = \widehat{cov}(\hat u_i, T_i) = 0$ and $\hat E \hat u_i = 0$.

~~Then, cov~~

$$\widehat{cov}(\hat u_i, R_i) =_1 \widehat{cov}(\hat u_i, 1-c_i-T_i)$$
$$=_2 -\widehat{cov}(\hat u_i, c_i) - \widehat{cov}(\hat u_i, T_i)$$
$$=_3 0$$

where $=_1$ follows by construction of c_i, T_i, R_i, $=_2$ follows by linearity of covariance, and $=_3$ follows by construction of the OLS regression.

ii ~~By construction of the OLS regressions,~~
~~$\hat\beta_x = \widehat{cov}(w_i, x_i)/\widehat{var}(x_i)$~~

By substitution of $R_i = 1 - c_i - T_i$ into ~~(2)~~ the OLS counterpart of (2),

$$w_i = \hat\gamma_0 + \hat\gamma_x x_i + \hat\gamma_c c_i + \hat\gamma_R(1-c_i-T_i) + \hat v_i$$
$$\cancel{= \gamma_0 + \gamma_x x_i +}$$
$$= (\hat\gamma_0 + \hat\gamma_R) + \hat\gamma_x x_i + (\hat\gamma_c - \hat\gamma_R)c_i - \hat\gamma_R T_i + \hat v_i$$

~~where~~

~~The corresponding OLS regression is~~

$$\cancel{w_i = (\hat\gamma_0 + \hat\gamma_R) + \hat\gamma_x x_i + (\hat\gamma_c - \hat\gamma_R)c_i - \hat\gamma_R T_i + \hat v_i}$$

where $\hat E \hat v_i = \widehat{cov}(x_i, \hat v_i) = \widehat{cov}(c_i, \hat v_i) = \widehat{cov}(R_i, \hat v_i) = 0$ by construction of the OLS counterpart of (2), hence $\widehat{cov}(T_i, \hat v_i) = \widehat{cov}(1 - c_i - R_i, \hat v_i) = 0$ by linearity of covariance.

Then the above equation obtained by substitution is an OLS regression of w_i on x_i, c_i, T_i. OLS regression uniquely decomposes w_i into a linear function of x_i, c_i, T_i, so the parameters of this regression coincide with the parameters of the OLS counterpart of (1).

$\hat\beta_x = \hat\gamma_x$, $\hat\beta_c = \hat\gamma_c - \hat\gamma_R$, $\hat\beta_T = -\hat\gamma_R$

$\hat\gamma_x = \hat\beta_x$, $\hat\gamma_c = \hat\beta_c + \hat\gamma_R = \hat\beta_c - \hat\beta_T$, $\hat\gamma_R = -\hat\beta_T$

This makes intuitive sense. $\hat\beta_x$ and $\hat\gamma_x$ each estimate the relationship between x_i and w_i in the population, so the two are equal. $\hat\beta_c$ estimates the relationship between ~~being in~~ living in a city and not in a rural area and wage, so it is equal to $\hat\gamma_c - \hat\gamma_R$, ~~the~~ where $\hat\gamma_c$ estimates the relationship between living in a city ~~rather th~~ and not in a town and wage and $\hat\gamma_R$ estimates the relationship between living in a rural area and not a town and wage. Similarly for $\hat\beta_T, \hat\gamma_c, \hat\gamma_R$.

b Estimate the population linear regression model

$$w_i = \delta_0 + \delta_x x_i + \delta_{cx} c_i x_i + \delta_{Tx} T_i x_i + \delta_c c_i + \delta_T T_i + \varepsilon_i$$

by OLS regression of w_i on x_i, $c_i x_i$, $T_i x_i$, c_i, T_i.

Test the hypothesis

$H_0: \delta_{cx} = \delta_{Tx} = 0$ against

$H_1: \delta_{cx} \neq 0$ or $\delta_{Tx} \neq 0$

with a F test, where the unrestricted model is the model above and the restricted model is

$$w_i = \delta_0' + \delta_x' x_i + \delta_c' c_i + \delta_T' T_i + \varepsilon_i'$$

Compute the ~~SSR~~ sum of squared residuals
$SSR = \sum_{i=1}^n \hat\varepsilon_i^2$, $SSR' = \sum_{i=1}^n \hat\varepsilon_i'^2$

Compute the F statistic
$F = \frac{n-k-1}{q} \frac{SSR' - SSR}{SSR}$
where $k = 5$, $q = 2$

Under the null, F is distributed according to the $F_{2,\infty}$ distribution.

Compare ~~the~~ F against a critical value c_α from this distribution at a suitable level of confidence α. Reject H_0 if $F > c_\alpha$.

If the null is rejected, conclude that $\delta_{cx} \neq 0$ or $\delta_{Tx} \neq 0$ hence the return to experience in a city $\delta_x + \delta_{cx}$ and/or the return to experience in a town $\delta_x + \delta_{Tx}$ is not equal to the return to experience in a rural area δ_x.

Date

4a H_0: ~~$\mu_A = \mu_B$~~
$\delta = 0$
H_1: $\delta \neq 0$
where $\delta = \mu_A - \mu_B$ where μ_A is the population mean monthly earnings in £ of trained employees and μ_B is that of untrained employees.

standard error
$se(\hat{\delta}) = \cancel{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}$ $\sqrt{s_A^2/n_A + s_B^2/n_B}$
$= \sqrt{(750)^2/100 + (750)^2/900}$
$= 79.057$

~~where σ_A and σ_B denote the respective~~
where s_A and s_B denote the respective sample standard deviations and n_A and n_B denote the respective sample sizes.

t statistic
~~$t(\hat{\delta})$~~ $t = \hat{\delta} - \delta_0 / se(\hat{\delta})$
$= (3800 - 3200) - 0 / 79.057$
$= 7.5895$

under the null, by CLT, given a sufficiently large sample, supposing that the distribution of earnings within the A and B groups is iid,
$t \xrightarrow{d} Z \sim N(0,1)$

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

For a two-sided test, $\alpha = 2\Phi(-c_\alpha) \Rightarrow c_\alpha = 1.960$

Reject the null. Conclude that mean earnings of trained employees are different from the mean earnings of untrained employees.

b Consider the causal model
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
where Y denotes earnings, X_1 is a dummy variable that takes value 1 iff a worker completed training, ~~and~~ X_2 is a variable that collects worker characteristics that dispose a worker to participate in training, and u collects the remaining unobserved determinants of Y.

Consider the population linear regression
$Y = \gamma_0 + \gamma_1 X_1 + v$

By construction,
$\gamma_1 = cov(Y, X_1)/var(X_1)$
$= cov(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, X_1)/var(X_1)$
$= [\beta_1 var(X_1) + \beta_2 cov(X_2, X_1)]/var(X_1)$
$= \beta_1 + \beta_2 cov(X_2, X_1)/var(X_1)$
$= \beta_1 + \beta_2 \pi_1$

where π_1 is the coefficient on X_1 in a population linear regression of ~~β_2~~ X_2 on X_1

The ~~population linear regression~~ coefficient on X_1 in a population linear regression of Y on X_1 alone decomposes ~~it to~~ into the causal effect of X_1 on Y, β_1, and a selection bias term.

This selection bias term is the product of the causal effect of X_2 on Y and the coefficient ~~of~~ on X_1 in a population linear regression of X_2 on X_1.

The selection bias term is likely to be positive because β_2 is likely to be positive and π_1 is likely to be positive. More motivated workers are more likely to participate in training (hence positive π_1) and are also likely to have higher wage (hence positive β_2). This is a problem in the current setting because the difference in average earnings is consistent for the population parameter γ_1 which overestimates the causal effect of interest β_1. Then, it is not possible to ~~conclude~~ infer the value (or even the sign of β_1).

Selection bias is likely to be small iff β_2 is small (which is unrealistic) or iff π_1 is small which is iff X_1 is approximately uncorrelated with X_2, which is if ~~X_1~~ X_1 is successfully randomly assigned.

c Supposing that training X_1 is successfully randomly assigned, it is uncorrelated with unobserved determinants of Y, including X_2, the ~~determinants of Y that ordinari~~ worker characteristics that dispose a worker to participate in training, which cause the selection bias. Then, ~~π_1~~ orthogonality holds in the ~~popula~~ short causal model ~~π~~ $Y = \beta_0 + \beta_1 X + u'$, where $u' = \beta_2 X_2 + u$, and the causal model is a population linear regression, whose parameters, in particular β_1, are consistently estimated by OLS regression of Y on X alone. In the context of binary treatment X_1, this estimator is the difference of sample means.

d Given that treatment is successfully ~~rand~~ randomly assigned, orthogonality holds in the short causal model and its parameters are consistently estimated by OLS regression, in particular the causal effect of interest is consistently estimated by the difference of means.

$H_0: \delta = 0$
$H_1: \delta \neq 0$
where $\delta = \mu_T - \mu_C$

$se(\hat{\delta}) = \sqrt{S_T^2/n_T + S_C^2/n_C} = \sqrt{750^2/50 + 750^2/50} = 122.47$

t-statistic
$t = \hat{\delta} - \delta_0 / se(\hat{\delta}) = (3625 - 3400) - 0 / 122.47 = 1.8372$

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

Reject the null if $t > c_\alpha$, where c_α is ~~the~~ the critical value from the standard normal distribution ~~for~~ at the $\alpha = 0.05$ level of significance.

For a two-sided test, $\alpha = 2\Phi(-c_\alpha) \Rightarrow c_\alpha = 1.960$

Fail to reject the null. ~~Fail to conclude~~ Conclude that employees do not have significantly different earnings at the 5% level of significance.

e Given that treatment is successfully randomly assigned, it is uncorrelated with unobserved determinants of wage such as employee motivation. Then, orthogonality holds in the "short" causal model, it is a population linear regression ~~on~~ of Y on X_1, hence its parameters are consistently estimated by OLS. In particular, the sample difference in means which coincides with the OLS ~~estimator for the~~ coefficient on X_1 is consistent for the corresponding population coefficient, which ~~is~~ coincides with the causal effect of interest β_1.

$\hat{\delta} = 3625 - 3400 = 225$ is consistent for the causal effect of interest.

f The hypothesis test is identical to that above except in the following:

$se(\hat{\delta}) = \sqrt{S_T^2/n_T + S_C^2/n_C} = \sqrt{750^2/80 + 750^2/20} = 108.25$
$t = (3625 - 3400) - 0 / 108.25 = 2.0785$

Reject the null. Conclude that employees have significantly different earnings at the 5% level of significance.

In general, where ~~sample~~ sample standard deviation is equal between the treatment and control groups, a ~~tes~~ RCT with comparably sized treatment and control groups is more powerful. ~~This follows from the formula for standard~~ Such tests have lower standard error hence larger t statistics.

g By inspection of the formula of the t-statistic, t is ~~##~~ a maximum when se is a minimum.

$se(\hat{\delta}) = \sqrt{750^2/n_T + 750^2/n_C}$
$= \sqrt{750^2/n_T + 750^2/(n - n_T)}$

$se(\hat{\delta})$ is a minimum when ~~$\hat{\delta}$~~ $se(\hat{\delta})^2$ is a minimum.

$se(\hat{\delta})^2 = 750^2/n_T + 750^2/(n - n_T)$

FOC: $(-1)750^2 n_T^{-2} + (-1)750^2(n - n_T)^{-2}(-1) = 0 \iff$
~~$750^2 [n_T^{-2}$~~
$n_T^{-2} = (n - n_T)^{-2} \iff$
$n_T = n - n_T \iff$
$n_T = n_C$

SOC: $(-1)(-2)750^2 n_T^{-3} + (-1)(-2)750^2(n - n_T)^{-3}(-1)(-1)$
$= 2(750^2)n_T^{-3} + 2(750^2)(n - n_T)^{-3}$
> 0 at $n_T = n - n_T > 0$

$se(\hat{\delta})^2$ is a minimum, $se(\hat{\delta})$ is a minimum and t is a maximum, given fixed ~~$\hat{\delta}$~~ difference in sample means and equal sample standard deviations when $n_T = n_C$.

h On the basis of the result in (e), we predict that the programme will yield an ~~£~~ increase in average earnings by £225, supposing that the ~~staff~~ employees in the two firms have a identical distribution of wage determinants, ~~whatever~~ ~~#~~ ~~treatment~~ ~~the effect~~ one employee's being trained has no effect on the value of training for another employee, i.e. treatment is individualistic and there are no spillover effects.

5a ~~E[Y|X]~~

$\cancel{EY = \beta_0 + \beta_1}$

$$E[Y|X] =_1 E[\beta_0 + \beta_1 X + \beta_2 X^2 + u \mid X]$$
$$=_2 \beta_0 + \beta_1 X + \beta_2 X^2$$

where $=_1$ follows by substitution, $=_2$ follows ~~from~~ by conditioning and given that $E[u|X]=0$

$$E[v|X] =_1 \cancel{E[Y} E[Y - \gamma_0 - \gamma_1 X \mid X]$$
$$=_2 E[Y|X] - E[\gamma_0 + \gamma_1 X \mid X]$$
$$=_3 \beta_0 + \beta_1 X + \beta_2 X^2 - \gamma_0 - \gamma_1 X$$
$$=_4 (\beta_0 - \gamma_0) + (\beta_1 - \gamma_1) X + \beta_2 X^2$$

where $=_1$ follows by substitution, $=_2$ follows by linearity of conditional expectation, $=_3$ follows by substitution and conditioning, $=_4$ follows by basic algebra.

$E[v|X] = 0$ iff $\beta_0 - \gamma_0 = 0$, $\beta_1 - \gamma_1 = 0$, and $\beta_2 X^2 = 0$. $E[v|X] = 0$ only if $\beta_2 X^2 = 0$. It is natural to suppose that this is implied by the question, so it is not possible that $E[v|X]=0$.

The given model implies that Y is constituted by a quadratic ~~deterministic comp~~ (in X) ~~+~~ deterministic component and some mean-zero ~~error~~ a stochastic error, ~~then the conditional expectation of Y is simply~~ that is mean independent of X so the conditional expectation of Y is simply this former term. The population linear regression gives the best linear (in X) prediction of Y. Graphically, the magnitude and direction of the error will be correlated with X.

b A and B obtain similar estimates for the coefficients in (3) with small standard errors. Their estimates are in general only few standard errors apart. This is because orthogonality holds for the causal model ~~#~~ (3) so OLS estimators are consistent for the ~~common~~ parameters of the causal model, which generates both sets of data.

A and B obtain dissimilar estimates for the coefficients in (4), with large standard errors.

One explanation for the dissimilar estimates is that A and B draw ~~#~~ data from populations ~~in which the range of X are X ranges has~~ with different ranges for X. By graphical intuition, the "best" linear approximation of a roughly quadratic distribution varies dramatically depending on ~~which ever~~ the domain of the approximation. A seems to have drawn data from a population in which X is generally higher.

The standard errors for both parameters for both A and B are large because ~~the~~ model (4) underfits the data and ~~is unable to capture the~~ so yields large residuals with high variance. This generally reduces the precision of OLS estimators.

c $\partial Y / \partial X \mid_{X=x} = \beta_1 + 2\beta_2 x$
$\partial Y / \partial X \mid_{X=x} = \gamma_1$

$\beta_1 + 2\beta_2 x = \gamma_1 \Leftrightarrow$

$\cancel{\operatorname{cov}(Y,X)/\operatorname{var}(X) + 2X\operatorname{cov}(Y,X^2)/\operatorname{var}(X) =}$

$\cancel{\operatorname{cov}(Y,\tilde{X})/\operatorname{var}(\tilde{X}) + 2x\operatorname{cov}(Y,\tilde{X}^2)/\operatorname{var}(\tilde{X})}$

$\cancel{= \operatorname{cov}(Y,X)/\operatorname{var}(X) \Leftrightarrow}$

~~cov(~~

$$\beta_1 + 2\beta_2 x = \operatorname{cov}(Y,X)/\operatorname{var}(X) \Leftrightarrow$$
$$\beta_1 + 2\beta_2 x = \operatorname{cov}(\beta_0 + \beta_1 X + \beta_2 X^2 + u, X)/\operatorname{var}(X) \Leftrightarrow$$
$$\beta_1 + 2\beta_2 x = \beta_1 \operatorname{cov}(X,X) + \beta_2 \operatorname{cov}(X^2, X)/\operatorname{var}(X) \Leftrightarrow$$
$$\beta_1 + 2\beta_2 x = \beta_1 + \beta_2 \operatorname{cov}(X^2,X)/\operatorname{var}(X) \Leftrightarrow$$
$$x = \operatorname{cov}(X^2,X)/2\operatorname{var}(X)$$

The two marginal effects agree where $x = \operatorname{cov}(X^2, X)/2\operatorname{var}(X)$.

The linear model underfits the causal model while the fifth order polynomial model overfits the causal model.

~~underfit~~ An underfitted model performs poorly on both training and test data because of high bias. Such a model has high bias because it does not have adequate flexibility to match the ~~underlying~~ process that generated the data. An underfitted model has low variance because fewer parameters are estimated with a finite amount of data and the model is less flexible, so less sensitive to noise in the training data. Then, in general, its performance on test data is poor but ~~has~~ relatively invariant to sampling variation.

An overfitted model performs well on training data but poorly on test data, because of high variance. Such a model has high variance because it ~~is~~ estimates a larger number of parameters with finite data and is highly flexible so highly sensitive to noise. In this case, a small amount of noise results in non-zero estimates of the coefficient on the fifth power, ~~which~~ and this term dominates for large X, ~~yielding~~ so the model has a large error for such X.

for the purpose of estimating the relationship between Y and X, provided that the range of X in the training data is sufficiently close to the range of X in which the relationship is of interest, this problem is not particularly large. Underfitting always yields poor estimates, even within the range of training data. Then, in such cases, overfitting is preferable.

e $E[Y|Z] =_1 E[\beta_0+\beta_1 X+\beta_2 X^2+u|Z]$

$=_2 \beta_0 + \beta_1 E[X|Z] + \beta_2 E[X^2|Z] + E[u|Z]$

$=_3 \beta_0+\beta_1(\pi_0+\pi_1 Z) + \beta_2(\pi_0+\pi_1 Z)^2$

$=_4 (\beta_0+\beta_1\pi_0+\beta_2\pi_0^2)$

$+(\beta_1\pi_1 + \beta_2 2\pi_0\pi_1)Z$

$+ \beta_2\pi_1^2 Z^2$

where $=_1$ follows by substitution, $=_2$ follows by linearity of conditional expectation, given $E[u|Z]=0$, $=_3$ follows by conditioning, given $E[\varepsilon|Z]=0$, $=_4$ follows by basic algebra.

$E[X|Z] = E[\pi_0+\pi_1 Z+\varepsilon|Z]$

$= \pi_0+\pi_1 Z$

by substitution, linearity of conditional expectation, and conditioning.

Let $\gamma_0 = \beta_0+\beta_1\pi_0+\beta_2\pi_0^2$, $\gamma_1 = \beta_1\pi_1+2\beta_2\pi_0\pi_1$, and $\gamma_2 = \beta_2\pi_1^2$.

$Y = \gamma_0+\gamma_1 Z+\gamma_2 Z^2+\xi$

is a population linear regression model of Y on Z and Z^2 because

$E[\xi|Z,Z^2] =_1 E[Y-\gamma_0-\gamma_1 Z-\gamma_2 Z^2|Z,Z^2]$

$=_2 E[Y|Z,Z^2] - E[\gamma_0+\gamma_1 Z+\gamma_2 Z^2|Z,Z^2]$

$=_3 E[Y|Z] - (\gamma_0+\gamma_1 Z+\gamma_2 Z^2)$

$=_4 0$

where $=_1$ follows by substitution, $=_2$ follows by linearity of conditional expectation, $=_3$ follows by conditioning, and $=_4$ follows by the result above, hence $cov(Z,\xi) = cov(Z^2,\xi) = E\xi = 0$.

Then population linear regression of Y on Z, Z^2 yields the parameters $\gamma_0, \gamma_1, \gamma_2$.

If $\pi_0=0$, then β_1, β_2 are recoverable from this regression. $\beta_1 = \gamma_1/\pi_1$, $\beta_2 = \gamma_2/\pi_1^2$.

7a age is included as a proxy for work experience ~~and hied is included as a control~~, ~~the~~ to control for the effect of work experience on wage. hied is included as a control.

The inclusion of these controls improves the precision of the OLS estimator $\hat{\beta}_1$ for population regression coefficient β_1. Plausibly, some of the variation in wage is accounted for by age and hied. Then, including these regressors reduces the magnitude and variance of OLS residuals. Supposing that heteroskedasticity is plausible, $se(\hat{\beta}_1) = n^{-1/2} \, var(\hat{u})/var(\cancel{x}\, alc)$. Then, a reduction in the variance of residuals reduces the ~~wage~~ standard error of $\hat{\beta}_1$ and increases the precision of this estimator.

The estimated coefficient on age cannot be given a causal interpretation because age is included as a proxy for work experience. ~~so the coefficient on age is~~ so it is presupposed that age has no direct causal effect on wage.

The coefficient on hied can be given a causal interpretation only if ~~it is~~ ~~exog~~ hied is ~~ex~~ exogenous, i.e. uncorrelated with the unmodelled determinants of wage. This is not ~~not~~ plausible. hied is likely correlated with intelligence and conscientiousness because a more intelligent and more conscientious person is more able to complete a university degree and these characteristics are determinants of wage.

The estimated coefficient on alc cannot be given a causal interpretation, similarly, because alc is endogenous due to omitted variable bias. Plausibly, alcohol consumption is correlated with number of hours worked (those who work less have more time to drink), which is a determinant of wage. Then, orthogonality fails in the causal model, it does not coincide with the population linear regression model, and OLS estimators are consistent for the parameters of the latter but not the former.

b A valid instrument is relevant, exogenous and excluded. An instrument is relevant iff it is plausibly correlated with the potentially endogenous regressor that it is an instrument for. An instrument is exogenous iff it is not correlated with unmodelled determinants of the dependent variable, i.e. it has no effect on the dependent variable through any variable other than that which it is an instrument for. An instrument is excluded iff it is not a direct causal determinant of the dependent variable. Then a valid instrument has an effect on the dependent variable only through the variable that it is an instrument for, and not directly nor through some other ~~variable~~ unmodelled variable.

hlth is plausibly relevant. It is known that high alcohol consumption is correlated with such health conditions. It is not plausibly excluded. Such health conditions are likely to undermine productivity hence have a negative effect on wage, especially in physically strenuous lines of work.

nchild is less plausibly relevant. Presumably a ~~m~~ man with more children has greater responsibilities and ~~is less likely~~ has fewer ~~opportun~~ occasions to drink. This relationship ~~is~~ does not seem particularly strong. nchild is not plausibly ~~excluded~~ exogenous. Plausibly, the more children a man has, the greater his financial responsibilities, hence the more likely he is to ~~find more lu~~ seek out more lucrative work and to be employed in more lucrative industries and/or roles.

psmoke is plausibly relevant. ~~P~~ Plausibly, a man whose parents smoked finds indulgence in the vices more normal and so is likely to consume more alcohol. psmoke is not plausibly exogenous. Men whose parents smoked are likely to be less healthy, having grown up in such an environment, and poor health plausibly undermines productivity, resulting in a lower wage.

None of the instruments are unambiguously valid. hlth is plausibly the most valid, ~~suppos~~ especially if the men in the study tend to work in less physically strenuous occupations.

c The estimates vary because alc is endogenous and ~~so~~ none of the controls is unambiguously valid, so each estimator is not consistent for the underlying causal effect of interest.

The coefficient in regression (i) likely has a negative bias because alc is positively correlated with negative determinants of wage such as ~~conscientousness~~ ~~and~~ ~~negatively correlated with post~~ impulsiveness and /or negatively correlated with positive determinants of wage such as number of hours worked.

The coefficient in (iv) ~~[illegible]~~ has a large negative bias. Plausibly, this is because psmoke is not a valid instrument and suffers from the same sort of endogeneity as alc. Men whose parents smoked, plausibly, grew up in less wealthy households and had ~~[illegible]~~ less access to educational ~~[illegible]~~ resources and professional networks through their parents. so ~~estimatt~~ the estimated coefficient on psmoke in a regression of wage on psmoke has a large negative omitted variable bias that ~~[illegible]~~ is inherited by the ~~[illegible]~~ ILS ~~coefficient~~ estimator for the coefficient on alc (which coincides with the 2SLS estimator where there is only one instrument).

Plausibly, the same is true to a smaller extent in (iii), ~~in general~~, ~~fewer~~ ~~men with higher~~ and ~~[illegible]~~ the estimated relationship between wage and nchild suffers from some negative omitted variable bias, which is inherited by the estimator for the coefficient on alc.

2.d The coefficient on alc in regression (iii) coincides with the coefficient obtained by ILS, given that there is only one ~~[illegible]~~ instrument. The coefficient obtained by ILS is equal to the coefficient on nchild in the reduced form regression divided by the coefficient on nchild in the first stage regression (vi). So the coefficient on nchild in the reduced form regression of lwage on nchild with controls age and hied is ~~equal~~ $-0.026 \times -5.71 = 0.14846$. This is a consistent estimator ~~of~~ for the causal effect of nchild on lwage iff nchild is exogenous (and the controls are exogenous) in this reduced form regression. ~~# Plausibly, there is no measurement error in nchild and nchild and lwage are not simultaneously determined, so this is consistent iff there is no omitted variable bias~~. Supposing further that the subsidies have no effect on lwage other than through nchild, the effect of ~~lwage on~~ the subsidies is an increase in lwage of $0.15 \times 0.14846 = 0.022269$.

Exogeneity of nchild in the reduced form regression is not plausible because nchild and lwage are likely to be simultaneously determined. A man with higher wage is more able to afford raising children and may choose to have more children as a result. So this estimate of the effect is not consistent.

ii Suppose that a reduction of the ~~[illegible]~~ proportion of smokers by 5 percentage points results in a reduction of the number of children ~~whose parents~~ who have at least one smoking parent by five percentage points ~~Suppose further that the public health programme in question is a counterfactual public health programme that would have had its effect on the parents of the men in this study Suppose further that the relationship between~~

Suppose further that regressors in regression (vii) are exogenous, then the estimated coefficient on psmoke in this regression is consistent for the causal effect of psmoke on alc. Suppose further that this causal effect is homogenous across the population (not just employed men aged 25-65), and that there are no spillover effects, so the result is externally valid. ~~Then the policy in question. Then let psmoke denote the sample mean of psmoke. A decrease in the proportion of smokers with families with smoking either parent smoking results in a decrease in the population mean of psmoke~~ The policy in question has a (estimated) causal effect of $-0.05 \times 7.43 = -0.3715$ on alc, supposing that it has no effect other than through psmoke.

Exogeneity in (vii) is implausible due to omitted variable bias. One such omitted variable is the sort of neighbourhood and/or community a person grows up in.

iii ~~Suppose that hith is a valid instrument (the problem of exclusion is insignificant because hith has only a small direct effect on lwage), then~~,

The coefficient on alc in (i) can be given a descriptive interpretation, even if it cannot be given a causal interpretation due to endogeneity. On average, a person ~~[illegible]~~ from the study population who consumed ~~[illegible]~~ 1 more unit of alcohol per week had lwage lower by 0.029, holding age and hied constant. So we expect the person who consumed 10 more units of alcohol per week to have lwage lower by 0.290. This supposes that the ~~causal effect of interest~~ relationship between alc and lwage is linear.

e $H_0: \beta_1 = 0$
$H_1: \beta_1 < 0$

t-statistic
$t = (\hat{\beta}_1 - \beta_0)/se(\hat{\beta}_1)$
$= \cancel{-0.0...}$
$-0.026/0.012$
$= -2.1667$

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

Then, the probability ~~of~~ under the null of observing a test statistic at least as unfavourable to the null as that actually observed is
$P(Z < -2.1667) = 0.015133$

p-value
$p = 0.015133$

Reject the null ~~of no change or effect~~ in favour of the alternative and conclude that alcohol consumption has a negative effect on earnings at all levels of significance $\alpha > p = 0.015133$.

Date

8a The process is non stationary. By inspection of the given figure, the fitted linear trend is downward-sloping. ~~The~~ By construction of the fitted linear trend, it is the best linear predictor of ~~the~~ SIE given t alone. Then, the expected value of SIE is ~~#~~ dependent on t, so SIE is not stationary.

By inspection, SIE appears to have a ~~linear deterministic~~ negative deterministic trend, a seasonality component, and a stochastic error component.

b SIE exhibits seasonality. By inspection, the value of SIE in any period is similar to ~~that~~ the value 12 periods (one year) prior. Each month series (is) appears to follow an ~~The given model effectively models the value of SIE in each period as a "union" of 12~~ approximately linear trajectory with some stochastic error.

The given model is a "union" of 12 ~~stochastic process~~ stochastic processes ~~time series~~, each corresponding to a given month, with a deterministic linear trend and a common (across the 12 time series) stochastic error term.

~~Treating the error term of each month stochastic series as having a common distribution is sensible if there is no a priori reason to think that the stochastic error term has a different distribution in each diff "season" (month). This condition fails if, for example, there are larger is reason to think that lower value lower values of SIE are associated with larger fluctuations because of "tipping points".~~

δ_i is some "initial" value of the deterministic component of ~~the~~ SIE in the ith month process that generated the observations in all ith months.

γ_i is the annual ~~the~~ change in the deterministic component of SIE in the ith month process, supposing that t is measured in years.

~~The model is not entirely plausible, the value of SIE in period t is likely to be dependent on the value of SIE in period t-1, and in particular, to be related to the stochastic error ε_{t-1}, # this is not accommodated in the given regression~~

c No ~~individual process that generate~~ time series corresponding to an individual month is stationary because the fitted trend lines are downward sloping, hence expected values decrease with increasing time, and are time dependent.

By inspection, the errors of the September regression are significantly larger than the errors of the March regression. It is not plausible to treat these errors as generated by a common process and having a common distribution. ~~The model should be revised to~~

$$\cancel{SIE_t = \sum_{i=1}^{12}(\delta_i D_{it} + \gamma_i D_{it} t + \varepsilon_i D_{it}\varepsilon_{it})}$$

This suggests that lower values of SIE are associated with larger fluctuations, plausibly ~~as~~ because of "tipping points" and instability around lower values.

Some of the fitted trend lines intersect. This is not plausible for such natural phenomena. We expect the pattern of seasonality to remain unchanged even in distant periods. Then, ~~this inters~~ that fitted trend lines intersect suggests ~~#~~ that (1) there is significant noise in the sample period resulting in ~~high~~ poor performance of the fitted model on ~~future~~ OOS periods because of high variance and (2) the model is not sufficiently flexible to capture the underlying dynamics of the ~~#~~ stochastic process ~~because ex~~ resulting in poor performance on ~~future~~ OOS periods because of high bias. In particular, the model fails to accommodate a plausible relationship between SIE_t and SIE_{t-1}, SIE_{t-2}, and so on.

d Compute heteroskedasticity-robust standard errors. By inspection, the magnitude of the regression residuals for the "March" regression ~~are~~ is smaller than the magnitude of the regression residuals for the "September regression". Formally, ε_t is larger (in magnitude) in ~~March~~ September periods than in March periods, ~~so so the assum~~ the former residuals have larger variance and (given that they are mean-zero by construction) ~~have to~~ their variance is equal to the expectation of their squares, so ε_t^2 is not time invariant, $E(\varepsilon_t^2|t) \neq E\varepsilon_t^2$ homoskedasticity is not plausible.

ii H_0: $\tilde{\gamma}_i = 0$ for all $i \in \{1, 2, \dots, 11\}$

H_1: $\tilde{\gamma}_i \neq 0$ for some $i \in \{1, 2, \dots, 11\}$.

Estimate the unrestricted model

$SIE_t = \delta_0 + \gamma_0 t + \Sigma_{i=1}^{11}(\tilde{\delta}_i D_{it} + \tilde{\gamma}_i D_{it} t) + \varepsilon_t$

by OLS

Estimate the restricted model

$SIE_t = \delta_0' + \gamma_0' t + \Sigma_{i=1}^{11}(\tilde{\delta}_i' D_{it}) + \varepsilon_t'$

by OLS

Compute the sum of squared residuals

$SSR = \Sigma_{t=0}^{T} \varepsilon_t^2$, $SSR' = \Sigma_{t=0}^{T} \varepsilon_t'^2$

Compute the F statistic

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

$= \frac{T - 23 - 1}{11} \frac{SSR' - SSR}{SSR}$

Under the null, $F \sim F_{q,\infty} = F_{11,\infty}$

Reject H_0 if $F > c_\alpha$, where c_α is the critical value drawn from the $F_{11,\infty}$ distribution at the $\alpha = 0.05$ level of significance.

$c_\alpha = 1.79$

If the null is rejected, conclude that there is no seasonal variation in the trend of SIE.

9a On average, a higher value of $CO_{2,t}$ (in ppm) by 1 is associated with a higher value of $Temp_t$ by 0.0080 (°C), holding $Temp_{t-1}$, Ec_t, $Ecob_t$, $Ecob_{t-1}$, and $Ecpr_t$ constant. Then, an increase in ~~CO2~~ CO_{2t} by 100 (ppm) ~~CO2~~ is associated with a 0.8000 (°C) increase in $Temp_t$.

The 95% confidence interval for the coefficient on CO_{2t} in the ~~above~~ given regression is

$C = [0.0080 - c_\alpha(0.0026), 0.0080 + c_\alpha(0.0026)]$
where ~~c_α~~ $\alpha = 0.05 = 2\Phi(-c_\alpha) \Rightarrow c_\alpha = 1.96$
$C = [0.002904, 0.013096]$

The 95% confidence interval for the immediate impact of an increase in ~~CO2~~ CO_{2t} by 100 (ppm) is

$100C = [0.29040, 1.3096]$

This random interval contains the true value of the immediate impact with 95% probability.

b The estimated ~~impact~~ immediate impact ~~of a~~ is given above. A permanent increase in CO2 in all periods $t, t+1, \ldots$ by 100 (ppm) has a ~~direct impact~~ immediate impact (calculated above) on Temp in the same period, but also an indirect impact ~~intermediated by~~ through the first lag of Temp.

Suppose that the permanent increase ~~beg~~ occurs at t.

$CO_{2t} \uparrow 100$ (ppm) $\Rightarrow Temp_t \uparrow 0.8000$ (°C)

$CO_{2t+1} \uparrow 100$ (ppm) $\wedge$ $Temp_t \uparrow 0.8000$ (°C)
$\Rightarrow Temp_{t+1} \uparrow 0.8000 + 0.879(0.8000) = 1.5032$ (°C)

$CO_{2t+2} \uparrow 100$ (ppm) $\wedge$ $Temp_{t+1} \uparrow 1.5032$ (°C)
$\Rightarrow Temp_{t+2} \uparrow 0.8000 + 0.879(1.5032) = 2.1213$ (°C)

In general, in period $t+\Delta t$, the total effect of a permanent increase in CO2 by 100 (ppm) is
$0.8000\left(\sum_{i=0}^{\Delta t} 0.879^i\right) = 0.8000\left(\frac{1-0.879^{\Delta t}}{1-0.879}\right)$

As Δt becomes large, this converges to
$0.8000\left(\frac{1}{1-0.879}\right) = 6.6116$ (°C)

c From (8), the coefficients on the orbital variables are ~~not~~ non-zero and significantly so (i.e. a large number of standard deviations from 0), then, CO_{2t} is likely correlated with these orbital variables.

From (7), the coefficients on the orbital variables are non-zero and significantly so. Then, ~~these variables are determinants of~~ presumably, these variables are determinants of Temp. (It is not plausible that these variables are proxies).

If the orbital variables are excluded in a "short" version of (7), then in this regression, ~~a~~ CO_{2t} is correlated with unmodelled determinants of $Temp_t$, then CO_{2t} is endogenous because of omitted variable bias. ~~Orthogonality fails in the short causal~~ Orthogonality fails in the short causal model that excludes the orbital variables, so it fails to coincide with the ~~popula~~ short population linear regression. OLS estimators are consistent ~~a~~ for the parameters of the latter but not the former. Then, the coefficients obtained by OLS in (7) cannot be given a causal interpretation, and the impact of anthropogenic ~~CO2~~ CO2 emissions on temperature in the Atlantic cannot be estimated.

d The onset of the agricultural revolution is estimated to be at ~10,000 BC. This ~~is~~ approximately ~~equal to~~ corresponds to $t=-12$.

~~Stat~~ Prior to approximately $t=-12$, the actual values of Temp and CO2 remained close to the dynamic forecasts and well within the 95% error band. Beginning around approximately $t=-12$, the actual values ~~begin~~ + diverge from the dynamic forecasts, and the actual value of CO2 ~~in to in t=-1 lies above~~ beginning around $t=-1$ ~~exceeds the~~ lies above the 95% error band.

This suggests that the process generating the observed data from $t=-100$ to $t=-12$ is the same process that generated observations from $t=-800$ to $t=-100$, but the process generating observations from $t=-12$ onward is a different process. There has been a structural break at $t=-12$, a new epoch begins around this time, which coincides with the beginning of the agricultural revolution.

The data offers some evidence of a human impact on the environment since the onset of the agricultural revolution. This is not decisive because the suspected structural break may well have been induced by a contemporaneous shock.

e the models treat the impact of CO_2 on ~~temp~~ Temp as linear. This is a ~~rea~~ potentially a reasonable ~~assum~~ approximation for small ~~changes~~ variations in CO_2, but for large ~~chang~~ variations, ~~#~~ there is no a priori reason to think this is plausible. There may well be "tipping points", "diminishing returns" or other sources of nonlinearity in the effect of CO_2 on temp.

The models are estimated on data from the past 800,000 years. It is not plausible that ~~this this data can~~ the process that generated this data is the same process that generates future observations ~~of~~ because a structural break ~~in the~~ due to the industrial revolution and other human impacts ~~of~~ on the ~~environ~~ environment other than CO_2 emission.