

ECON 545 - 2014

Solution for Final Exam

Q. 1. (Briefly)

(a) Std. errors will be inconsistent - need to use Newey-West correction. F & t tests will be invalid. OLS will be inefficient.

(b) R^2 & F are (+ve) monotonically related.
When one rises, the other one must rise too.

(c) Testing from "General" to "Specific" is preferred. Less likely to omit & relevant regressor. This is preferable to including irrelevant regressors. (Mention, bias, efficiency, consistency.)

(d) Nothing wrong with this. White's covariance estimator is still consistent even if the errors are heteroskedastic.

(e) I.V. estimator will be consistent. Usually biased in finite samples.

Q.1

$$\begin{aligned}
 \text{(a)} \quad \hat{\beta}_1 &= (x_1' \Omega^{-1} x_1)^{-1} x_1' \Omega^{-1} y \\
 &= (x_1' \Omega^{-1} x_1)^{-1} x_1' \Omega^{-1} [x_1 \beta_1 + x_2 \beta_2 + \varepsilon] \\
 &= \beta_1 + (x_1' \Omega^{-1} x_1)^{-1} x_1' \Omega^{-1} x_2 \beta_2 \\
 &\quad + (x_1' \Omega^{-1} x_1)^{-1} x_1' \Omega^{-1} \varepsilon
 \end{aligned}$$

$$\Rightarrow E(\hat{\beta}_1) = \beta_1 + (x_1' \Omega^{-1} x_1)^{-1} x_1' \Omega^{-1} x_2 \beta_2,$$

$$\text{as } E(\varepsilon) = 0.$$

So, $E(\hat{\beta}_1) \neq \beta_1$, unless $x_2 \beta_2 = 0$ (which just means that (4.2) is the true model), or more generally, unless $(x_1' \Omega^{-1} x_2) = 0$.

(b) We have applied GLS to a model that omits some relevant regressors. That is, we have applied GLS with an invalid set of (linear) restrictions — namely, $\beta_2 = 0$. In the case of OLS this will reduce the variability of the estimator, even though the restrictions are false. Now, GLS is just OLS applied to the transformed data, $y^* \& x^*$, where $y^* = Py$, $x^* = Px$, & $P'P = \Omega^{-1}$. So, the same result will apply here. The variability of the GLS estimator of β_1 in (4.2) will be less than that of the GLS estimator of β_1 in (4.1).

25

(c) We estimate the model:

$$\begin{aligned} y &= X_1 \beta_1 + X_2 \beta_2 + u \\ &= (X_1, X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + u \\ &= X\beta + u. \end{aligned}$$

The GLS estimator will be:

$$\begin{aligned} \hat{\beta}_G &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \\ &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} (X\beta + \varepsilon) \end{aligned}$$

Now, noting that $X_1 = (X_1, X_2) \begin{pmatrix} I \\ 0 \end{pmatrix} = XS$, we have:

$$\hat{\beta}_G = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} (XS\beta + \varepsilon)$$

$$= S\beta + (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \varepsilon$$

$$\therefore E(\hat{\beta}_G) = S\beta + 0 ; \text{ as } E(\varepsilon) = 0$$

$$= \begin{pmatrix} I \\ 0 \end{pmatrix} \beta = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}$$

$$\text{So, } E \left[\begin{array}{c} \hat{\beta}_G \\ \hat{\beta}_{2G} \end{array} \right] = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \text{ as } \beta_2 = 0.$$

So, the estimator is unbiased for both β_1 & β_2 .

(d) They will be consistent. Here we have failed to impose the (valid) restriction that $\rho_2 = 0$. In the OLS case this will lead to a loss of efficiency, relative to what could have been achieved by estimating (5.3) instead of (5.4). The same is true here, with GLS, as GLS is just OLS with the transformed data. So, our estimator in this case will be inefficient, both in finite samples & asymptotically.

Q.3.

(a) Discuss signs & significance of the estimated coefficients. Mention R^2 .

(b) Time-series data. To allow for possibility that autocorrelation in errors may render the std. errors inconsistent.

(c) In results 2, regular std. errors have been used. This results in loss of significance of 2 regressors (especially Trend). This suggests, loosely, that autocorrelation maybe present. However — note that $n = 26$ (small).

(d) Result 3 - H_0 : Serial Independence
 H_A : AR(1) / MA(1)

Reject H_0 ($p = 0.014$) — there is 1st-order

autocorrelation.

Result 4 - H_0 : Serial Independence
 H_A : AR(2) / MA(2)

Reject H_0 ($p = 0.0036$) - there is 2nd-order autocorrelation.

The tests don't use the std. errors, so it doesn't matter if we have used Newey-West or not.

(e) $\hat{E}_t = \rho_1 E_{t-1} + \rho_2 E_{t-2} + u_t + \phi_1 u_{t-2}$.

(f) t-test on dummy variable. $p = 0.0155$.

At the 5% sig. level, I'd reject H_0 that coefficient = 0. It had a significant effect.

(g) This variable, & the dependent variable, are both measured in logarithms, so the coefficient is an elasticity.

Interpretation: A 10% increase in average output of coal will result in an 8.1% increase in total injuries, ceteris paribus.

C.I. $0.080981 \pm t_{(5)} 0.016371$ (per unit)

or, $0.080981 \pm (2.571)(0.016371)$

or, $[0.0389 ; 0.1231]$.

Q.4

(a) See Lecture Slides.

$$(b) e^{*'} e^* = e'e + (Rb - q)' [R(R'x) - R']^{-1} (Rb - q).$$

$x'x$ is pos. def., & R has full rank,
so $R(R'x)R'$ is pos. def.. So, $[R(R'x) - R']^{-1}$
is positive definite.

This implies $(Rb - q)' [R(R'x) - R']^{-1} (Rb - q)$
is generally positive. It could be zero —
eg if $Rb = q$.

So, $e^{*'} e^* \geq e'e$.

(c) If $Rb = q$. That is, if the unrestricted
OLS estimator of β happened to already
satisfy the restrictions, $R\beta = q$.

$$(d) y_i = (1 - \beta_2) + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_3 x_{4i} + \varepsilon_i$$

$$\text{or, } (y_i - 1) = \beta_2 (x_{2i} - 1) + \beta_3 (x_{3i} + x_{4i}) + \varepsilon_i$$

$$\text{or, } g_i^* = \beta_2 x_{2i}^* + \beta_3 (x_{3i} + x_{4i}) + \varepsilon_i$$

(e) It will be unbiased if both restrictions
are correct. Otherwise it will be biased.
If both restrictions are correct it will be
more efficient than OLS. If only one

I

restriction is correct, or if both are wrong, it may have smaller or greater MSE than OLS.

Q.5. (This is the so-called "Mixed Regression Estimator" - Theil.)

(a)

$$\begin{aligned} y &= X\beta + \varepsilon \\ q &= R\beta + v \end{aligned}$$

or,

$$z = \begin{pmatrix} y \\ q \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix}\beta + \begin{pmatrix} \varepsilon \\ v \end{pmatrix}$$

(b)

$$\begin{aligned} \hat{\beta} &= \left[\begin{pmatrix} X \\ R \end{pmatrix}' \begin{pmatrix} X \\ R \end{pmatrix} \right]^{-1} \left[\begin{pmatrix} X \\ R \end{pmatrix}' \begin{pmatrix} y \\ q \end{pmatrix} \right] \\ &= [(X' R') \begin{pmatrix} X \\ X \end{pmatrix}]^{-1} [X' R'] \begin{pmatrix} y \\ q \end{pmatrix} \\ &= (X' X + R' R)^{-1} (X' y + R' q) \end{aligned}$$

$$(c) E(\hat{\beta}) = (X' X + R' R)^{-1} [X' E(y) + R' E(q)]$$

$$= (X' X + R' R)^{-1} (X' X \beta + R' R \beta)$$

$$= (X' X + R' R)^{-1} (X' X \beta + R' R \beta)$$

$$= \beta. \quad (\text{Unbiased})$$

(8)

$$\begin{aligned}
 \underline{\text{(d)}} \quad V(\hat{\beta}) &= (X'X + R'R)^{-1} V(X'y + R'q) (X'X + R'R)^{-1} \\
 &= (X'X + R'R)^{-1} \{ X'V(y)X + R'V(q)R \} (X'X + R'R)^{-1} \\
 &= (X'X + R'R)^{-1} (\sigma^2 X'X + R'VR) (X'X + R'R)^{-1}
 \end{aligned}$$

(e) If $V = \sigma^2 I$, then

$$\begin{aligned}
 V(\hat{\beta}) &= \sigma^2 (X'X + R'R)^{-1} (X'X + R'R) (X'X + R'R)^{-1} \\
 &= \sigma^2 (X'X + R'R)^{-1}
 \end{aligned}$$

$$\underline{\text{(f)}} \quad V(b) - V(\hat{\beta}) = \sigma^2 \{ (X'X)^{-1} - (X'X + R'R)^{-1} \}.$$

Definiteness of $V(b) - V(\hat{\beta})$ is the same

as definiteness of $(X'X)^{-1} - (X'X + R'R)^{-1}$,
which is same as definiteness of

$$(X'X + R'R) - (X'X), \text{ or of } R'R.$$

R has full rank, so $R'R$ is pos. def.,

& $\hat{\beta}$ is more efficient than b .

$$\underline{\text{(g)}} \quad \text{Use } t_{(n-k)} = \frac{\hat{\beta}_i - \beta_0}{\sqrt{s^2 (X'X + R'R)^{-1}_{ii}}}$$

$$\text{where } s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(n-k)$$

& $t_{(n-k)}$ will be Student-t, because
 $\hat{\beta}$ is normal and independent of s^2 .

Q.6 (a) Most of the coefficients have the anticipated signs: I'd expect that the longer the duration of the event, the greater the number of people who might observe it; the more reliable are the witnesses, the more likely that something actually was visible in the sky, and the greater the likely number of witnesses; and the outlier dummy should be positively related to WIT as it is unity when WIT is very large. The only surprise is the sign of STR - I'd expect that if a truly weird event occurred in the sky, this would result in more witnesses than if a "somewhat odd" event occurred.

However, STR is not statistically significant, anyway. The other regressors (excluding the intercept) are significant on 1-sided tests. The R^2 is quite low, but these are cross-section data, so this is really not surprising.

(b) H_0 : errors are homoskedastic
 H_A : errors are heteroskedastic.

This is only an asymptotically valid test - we need a large sample size. Here we have $n = 210$, so the test should be reasonably reliable. Looking at nR^2 , the p-value is 0.0478. I would reject H_0 at the 5% or 10% significance levels, and conclude that there is some form of heteroskedasticity.

(c) The model has been modified in terms of the regressors, but this nothing to do with the result of White's test. The outcome of the latter test has prompted us to use White's consistent estimator of the covariance matrix for $\hat{\beta}$, & hence consistent standard errors.

(d) We are using an asymptotically valid Wald test to test if the coefficients of the 3 dummy variable terms are all zero. The null is

$$H_0: \beta_4 = \beta_5 = \beta_6 = 0$$

& H_A : At least one of $\beta_4, \beta_5, \beta_6$ not zero.

The p-value for the χ^2 test is 0.1295, so I cannot reject H_0 at the 10% (or lower) level. I'd conclude that people in B.C. are no different from those in other provinces when it comes to UFO sightings!

(e) It suggests that there is heteroskedasticity, perhaps of the form $\sigma_i^2 \propto 1/\text{STR}$, or something similar to this.

(f) Weighted Least Squares has been used to estimate the model, modified by deleting the (insignificant) dummy variable terms. If I assume that $\sigma_i^2 \propto 1/\text{STR}$, then $\text{WEIGHT} = (\text{STR})^{1/2}$.

Q.7.

- (a) (i) \hat{b} will be unbiased, consistent, but inefficient. If one or more of the x 's are lagged values of y , \hat{b} will be inconsistent.
- (ii) These will be invalid, as the wrong estimator of $V(\hat{b})$ will be used to get the std. errors that are used to construct C.I.
- (iii) These statistics will not be t -distributed.

$$(b) \quad y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_K x_{Kt} + \epsilon_t$$

$$\rho y_{t-1} = \rho \beta_1 + \rho \beta_2 x_{2t-1} + \dots + \rho \beta_K x_{Kt-1} + \rho \epsilon_{t-1}$$

$$y_t = \beta_1(1-\rho) + \rho y_{t-1} + \beta_2 x_{2t} - \rho \beta_2 x_{2t-1} + \dots$$

$$- \dots + \beta_K x_{Kt} - \rho \beta_K x_{Kt-1} + u_t.$$

- (c) The model is now one with a "well-behaved" error term, but it is non-linear in the parameters.

- (d) Mention Newton-Raphson & various convergence issues. Need to get a global min. of the sum of squared residuals.

(e) Mention the Wald test. We want to test if $\rho = 0$, so this is just a t-test, & the statistic would be $N(0, 1)$ in very large samples. Only an asymptotic result

(f) $|p| < 1$, for stationarity. If not satisfied, the error process does not have a finite variance & we can't get sensible estimates of the model's parameters.