Mid-Term 4 Answers

1. Model in levels so estimated coefficients give the change in wages (in £), of a unit change in the level of the rhs variable.
   So 1 year of education raises wages by £0.50 (dW/dEDUC = 0.50)
   Experience is a quadratic which means the effect on wages is non-linear (not constant). Coefficients suggest wages rise then fall with experience

   dW/dEXP = 1.00-2*(.01)AGE

   All variables are statistically significantly different from zero.
   Regression explains around 25% of variation in dependent variable

   Wage maximised when dW/dEXP = 1.00-2(.01)AGE = 0 (f.o.c. for max)

   So 1.00=0.02EXP AGE=50

   Wages are maximised at age 50

b) F test of hypothesis that age coefficients should be included in model involves comparing RSS from unrestricted model (with age and age2) and restricted regression (without age and age2)

Either use

$$F = \frac{RSSrestrict - RSSunrestrict\ /J}{RSSunrestrict\ /N- Kunrestrict} \sim F(J, N-Kunrestrict)$$

1st find RSS in both equations using $R^2 = 1-RSS/TSS$ so $RSS = (1-R^2)TSS$

in (1) RSS = (1-.25)*100 = 75
in (2) RSS = (1-.2)*100 = 80

J = No. excluded variables = 2 (age, age squared)

so F becomes

$$= \frac{80 - 75\ /2}{75/104-4} \sim F(2, 104-4)$$

$$= 3.333$$

(or use alternative version of test

$$F = \frac{R^2unrestrict - R^2restrict\ /J}{1-R^2unrestrict\ /N- Kunrestrict} \sim F(J, N-Kunrestrict)$$

$$= \frac{0.25 - 0.20\ /2}{1-0.25/104-4} \sim F(2, 104-4)$$

$$= 3.333$$)

From F tables, critical value at 5% level F(2, 100) = 3.10
So estimated F > Fcritical

Reject null that excluded coefficients are zero (ie they should be in the model)

iii) Omitted variables are statistically significant so should have been included. Omission of relevant variables makes OLS estimates of coefficients and standard errors (and T and F tests) on remaining variables biased.

iv) Test is for functional form (omitted variables) is Ramsey RESET test

2. Measurement error in the dependent variable means OLS remains unbiased but standard error of estimates is larger

True: $Var(\hat{\beta}) = \dfrac{\sigma_u^2}{NVar(X)}$

Estimate: $Var(\tilde{\beta}) = \dfrac{\sigma_u^2 + \sigma_e^2}{NVar(X)}$

So true variance is 100/100*5 = 0.2
Estimated = 100+100/100*5 = 0.4

(Variance of estimate doubles)

a) true ols = cov(X^{true},y)/Var(X) =
Cov(Numcigs,$income^{true}$)/Var($Income^{true}$) = 5/5 = 1

b) observed OLS = Cov($X^{observed}$, y)/Var($X^{obsreved}$) = 1/5 = 0.2

c) so estimated value is biased toward zero (attenuation bias) in presence of measurement error in right hand side variable,

$$\hat{b}_1 = \frac{Cov(X^{observed}, y^{true})}{Var(X)} = b_1 + \frac{-b_1 Var(w)}{Var(X)} \neq b_1$$

(measurement error likely to increase the variance of X variable and reduce covariance between X variable and y variable. Both these factors work to reduce size of estimated OLS coefficient)

3 Using order condition for identification. Equation is identified if

K – k >= m – 1

Where K = total no. of exogenous variables in **system**
k = no. of exogenous variables in **any one equation**
m = no. of endogenous variables in **any one equation**

In (1) K = 2 (Wages, Income) k=2 (wages, income) m =2 (Price and output)
So 2– 2 < 2-1

(1) not identified (no exogenous variables in the system of equations that does not appear in (1) and so no instrument is available

In (2) K = 2 (wages, income) k = 0 m =2 (price and output)
So 2 – 0 > 2-1

(2) over identified (two exogenous variables in the system of equations that don't appear in (2))

so can use either wages or income as an instrument for price in (2)

We know that in large samples it is more efficient to use both

$$\hat{b}_{2SLS} = \frac{Cov(\hat{X}, y)}{Cov(\hat{X}, X)} = \frac{Cov(\hat{price}, output)}{Cov(\hat{price}, price)}$$

```
Wu-Hausman says regress badly measured variables on instrument
Save residuals and include as extra regressor in original equation.
Then t test on significance. If residuals are significantly different
from zero, conclude that OLS and IV estimates will be different.
There is endogeneity bias in OLS equation because of correlation of
badly measured experience variable with residual.
```

4.
See lecture notes. Hence OLS coefficients in column 1 will be unbiased but standard errors and therefore t and F values and confidence intervals are all invalid.
(Consequence: may mistakenly conclude variable is statistically significant from zero, for example, when it is not).

DW test indicates that regression suffers from autocorrelation
DW test critical values depends on number of observations and number of variables in regression (excluding constant)

So in this case T=200 (50*4=200) and k'=1
(this is **quarterly** data, so multiply number of years (50) by 4 to get total number of obs.)

From tables DWlow = 1.76, DWupper = 1.78

So estimated DW < DWlow, hence accept that there is 1st order (positive) autocorrelation present.

In (2) DW now > DW_{upper} so no autocorrelation on this basis. But we know DW biased toward 2 in presence of lagged dependent variable. So test invalid

With lagged dependent variables can either use Durbin h test or (asymptotically valid) Breusch-Godfrey test instead. However given

$$h = \hat{\rho}\sqrt{\frac{T}{1 - TVar(\hat{\lambda})}}$$

where T = sample size (number of time periods) and var(λ) is the estimated variance of the coefficient on the lagged dependent variable = .3*.3 (given standard error = .3 in the question). This means that

$$TVar(\hat{\lambda}) = 200 * .09 = 18$$

so that

$$1 - TVar(\hat{\lambda}) = 1 - 18 < 0$$

and so Durbin h cannot be calculated (involves square root of a negative number). Hence need to use Breusch-Godfrey.