Economics 421
Winter 2009
Final Exam

1 (a) Heteroskedasticity means that the variance of the error term is unequal across observations.

(b) Examples can be about learning over time → errors↓, the use of scale variables covering a broad range, e.g. sales or income from very small to very large, better data collection over time, outliers, or model misspecification. To be specific, a regression of sales on advertising expenditures might be heteroskedastic if Wal-Mart and the corner store are in the same data set (since the variance in the large firm's sales might be much larger than the small firms)

(c) To use White's test, perform the following steps:

- estimate the model by OLS, save the estimated residuals $\hat{u}_i$
- regress $\hat{u}$ on a constant, X_2, X_3, X_2^2, X_3^2 and X_2X_3.
- Compute $NR^2 \sim \chi^2(5)$
- Compare NR^2 to critical value for χ^2. If larger than critical value, reject, else fail to reject.

2\. (a) The consequences are

- OLS remains unbiased and consistent
- OLS no longer BLUE (not Best).
- if there are lagged dep variables, no longer unbiased/consistent
- Generally the estimates of $\hat{\sigma}_{\hat{\beta}}$ are biased. The bias in the estimate of the standard error $\rightarrow$ biased test statistics; e.g.

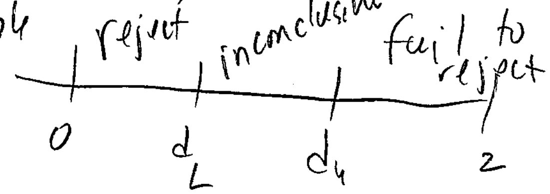
$$t = \frac{\hat{\beta} - \beta_{H_0}}{\hat{\sigma}_{\hat{\beta}}} \rightarrow \text{biased}$$

- note also that if there is positive serial correlation and the X^s are growing over time, $\hat{\sigma}_{\hat{\beta}}$ biased $\downarrow$ (t^s too big). [Vice-versa for neg corr]

(b) To conduct a DW test

- estimate the model by OLS, save $\hat{u}_t$
- Compute $dw = \dfrac{\sum_2^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_1^T (\hat{u}_t)^2}$. This is the DW stat.

- To test for pos correlation, look up d_u and d_L in a table



[cont.]

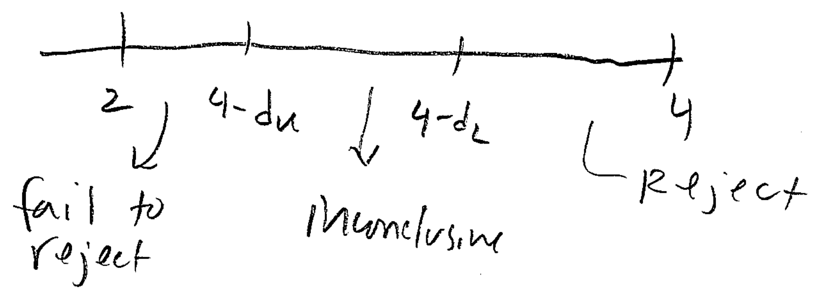
The values in the table depend upon T, the number of observations, and $K-1$, the # of coeffs minus the constant.

If dw is such that $0 < dw < d_L \Rightarrow$ reject

$d_L < dw < d_U \rightarrow$ inconclusive

$d_U < dw < 4 \rightarrow$ fail to reject

- For negative serial correlation, just flip it around



The test does not capture higher order serial correlation, it is designed to detect first order correlation only, i.e an error term of the form

$$u_t = \rho u_{t-1} + e_t, \quad |\rho| < 1$$

If, say, $u_t = \rho u_{t-4} + e_t$, $|\rho| < 1$, the test will miss the relationship.

3 (a) A valid IV must satisfy these properties

- it must be correlated with the X it is instrumenting for
- it must be uncorrelated with the error term
- it cannot itself be an explanatory variable

(b) The usual OLS estimator is

$$\hat{\beta}_2^{OLS} = \beta_2 + \frac{\Sigma(X_i - \bar{X})(u_i - \bar{u})}{\Sigma(X_i - \bar{X})^2}$$

and $\text{plim}\ \hat{\beta}_2^{OLS} = \beta_2 + \frac{\text{cov}(X,u)}{\text{var}(X)}$. If X and u are correlated, the estimator is inconsistent.

To overcome, use the IV estimator

$$\hat{\beta}_2^{IV} = \frac{\Sigma(Z_i - \bar{Z})(Y_i - \bar{Y})}{\Sigma(Z_i - \bar{Z})(X_i - \bar{X})}$$ where Z_i satisfies the properties in part (a)

Then $$\hat{\beta}_2^{IV} = \beta_2 + \frac{\Sigma(Z_i - \bar{Z})(u_i - \bar{u})}{\Sigma(Z_i - \bar{Z})(X_i - \bar{X})}$$

So that $$\text{plim}\ \hat{\beta}_2^{IV} = \beta_2 + \frac{\text{cov}(Z,u) \rightarrow \text{zero by the assumption in (a)}}{\text{cov}(Z,X) \rightarrow \text{non-zero by assumption in (a)}}$$

$$= \beta_2$$

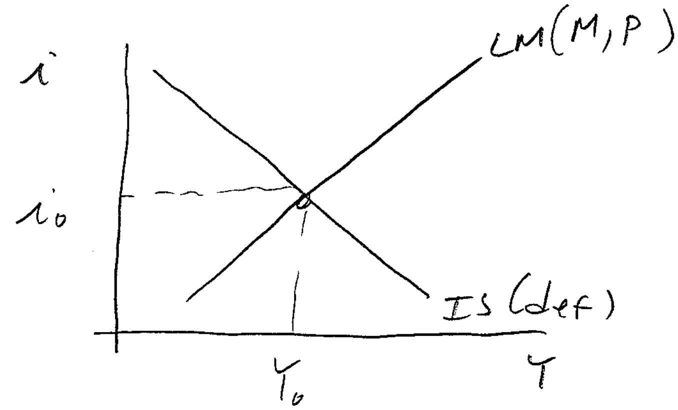
so, consistent

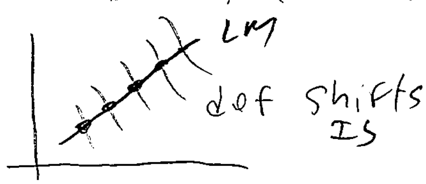
(c) To get the IV, regress each endogenous RHS variable on all the X^s (exog vars), then use $\hat{y}$ (the predicted value of the endog. variable) as the IV [correlated w/ y, uncorr w/ error].

(4) The model is

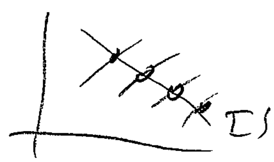
$$y_t = a - b\,i_t + c\,def_t + u_t \qquad IS$$
$$i_t = f + g\,Y_t - h\,M_t + K\,P_t + v_t \qquad LM$$

(a) As def moves up and down, the LM curve is mapped out $\Rightarrow$ def ID's the LM





As M moves, the IS is ID



But, As P moves, the IS is also ID. So, LM exactly IS (only one mapped out), but IS over ID (two different curves are mapped out, which to choose).

math: equation 1

two exclusions $\geq g-1 = 1$

$\Rightarrow$ over ID

equation 2

one exclusion $\geq g-1 = 1 \Rightarrow$ exact ID

[Cont.)

(4) [Cont.] (b)

$$y = a - b\{f + gY - hM + kP + v_t\} + c\,Def + u_t$$

$$Y + bgY = (a - bf) + bhM - bkP - bv_t + c\,Def + u_t$$

$$\longrightarrow \quad Y_t = \frac{1}{1+bg}\left[(a-bf) + bhM_t - bkP_t + c\,Def_t - bv_t + u_t\right]$$

$$i_t = f + g(a - bi + c\,def + u) - hM + kP + v$$

$$i + bgi = (f + ga) + gc\,def - hM + kP + v + gu$$

$$\longrightarrow \quad i_t = \frac{1}{1+bg}\left[(f+ga) + gc\,Def_t - hM_t + kP_t + v_t + gu_t\right]$$

(5) (a) perfect multicollinearity means there is an exact linear relationship among some of the X^s in the model, e.g.

$$2X_3 - 4X_2 + 3 = 0.$$

Imperfect multicoll. is an inexact version of this, e.g.

$$2X_3 - 4X_2 + 3 = V_t$$

, where V_t is a random variable.

[Cont.]

5 (a) [Cont.]

When there is perfect multicollinearity model OLS breaks down, estimates cannot be obtained. The solution in this case is to drop one of the variables (since it is entirely redundant in terms of providing any new information).

When it is imperfect, estimates can be obtained, but they are imprecise. OLS cannot tell the variables apart with any precision, and hence does not know what values to give to individual coefficients. The best solution is to collect more data, enough so that there is sufficient independent variation to tell the X^s apart. However, if the correlation is large, e.g. > 80%, a variable can be dropped. [Other options are to do nothing, transform variables (e.g. log)].

(b) Detection: One sign is a high R^2 and insignificant t-stats (or F-stat on group of variables signif ⇒ group matters, but indiv. t^s insignificant). Some drop a variable if the pairwise correlation > .8.

(6) (a) The linear probability model is

$Y_t = \alpha + \beta X_t + u_t$ (can have more than one X)

this is a probability, ∴ it is 0 or 1 representing the outcome [some choice for each indiv.]

Let $P_t = \text{Prob}(Y_t = 1)$

$1 - P_t = \text{Prob}(Y_t = 0)$

Then, using that $E(u_t) = 0$

$$0 = E(u_t) = P_t[1 - \alpha - \beta X_t] + (1 - P_t)(-\alpha - \beta X_t)$$

[variance formula for binomial is the sum of the probabilities times the outcomes]

from this we get that $P_t = \alpha + \beta X_t$

Next,

$$\sigma_t^2 = E(u_t^2) = P_t(1 - \alpha - \beta X_t)^2 + (1 - P_t)(-\alpha - \beta X_t)^2$$

sub from above

$$\sigma_t^2 = P_t(1 - P_t)^2 + (1 - P_t)(-P_t)^2$$
$$= P_t(1 - P_t)$$

$\sigma_t^2 = (\alpha + \beta X_t)(1 - \alpha - \beta X_t)$ which varies with each obs ⇒ heteroskedasticity

[cont.]

The correction is the usual one for heteroskedasticity

Use that $\hat{\sigma}_t^2 = \hat{Y}_t(1-\hat{Y}_t)$

$\hookrightarrow$ estimate of P_t

- estimate by OLS, save $\hat{y}_t$
- $\hat{\sigma}_t^2 = \hat{y}_t(1-\hat{y}_t)$
- divide variables in the model by $\hat{\sigma}_t$ to get "star" values
- regress y_t^* on $\frac{1}{\hat{\sigma}_t}$ and x_t^*

$y_t^* = \frac{Y_t}{\hat{\sigma}_t}$, $x_t^* = x_t/\hat{\sigma}_t$

(b) The Logit model is

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta x + u$$

where $0 \le p \le 1$ is a probability. This model is attractive because, unlike with the linear probability model,

$$P = \frac{1}{1+e^{-(\alpha+\beta x+u)}}$$ is <u>always</u>

between zero and one.

7 Maximum likelihood assumes a set of
(a) data are drawn from some distribution, e.g. a normal, with unknown parameters (e.g. the mean μ and the variance σ^2 that fully characterize a normal).

The estimator finds the parameters of the distribution that are most likely to have generated that observed data

(b) The properties are

- The estimates are consistent (but may be biased in small samples
- The estimates are asymptotically efficient. For large N, no other estimator has a smaller variance
- The estimates are asymptotically normal. The MLEs closely approx a normal even if the underlying dist. is non-normal.