Hetroskedasticity

consequence:	the statistics used to test hypotheses under Gauss-Markov assumptions are not valid in the presence of hetroskedasticity.
Valid estimator (any form)	∑[(x1- x-)^2 uˆi^2]/[SST^{2}x]
	SSTx=∑(x1- x-)^2
Robust Standard error	Varˆ(βˆj)=∑[rˆij^{2}û^{2}i]/[SSR^{2}j]

properties of OLS Matrix

Sum of Squared Residuals	(y − Xβˆ)′(y − Xβˆ)
	y′y − βˆ′X′y − y′Xβˆ + βˆ′X′Xβˆ
	y′y − 2βˆ′X′y + βˆ′X′Xβˆ
Minimise the SSR	∂(SSR)/∂βˆ = −2X′y + 2X′Xβˆ = 0
from the minimum we get: "normal equation"	(X′X)βˆ = X′y
Solve for OLS estimator βˆ; by pre multiplying both sides by (X′X)	(X′X)−1(X′X)βˆ = (X′X)−1X′y
by definition, (X′X)−1(X′X) = I	Iβˆ = (X′X)−1X′y
	βˆ = (X ′ X)−1 X ′ y
Properties	
The observed values of X are uncorrelated with the residuals.	X′e = 0 implies that for every column xk of X, x′ke = 0.
substitute in y = Xβˆ + e into normal equation	(X′X)βˆ = X′(Xβˆ + e)
	(X′X)βˆ = (X′X)βˆ + X′e
	X′e = 0

properties of OLS Matrix (cont)

The sum of the residuals is zero.	If there is a constant, then the first column in X (i.e. X1) will be a column of ones. This means that for the first element in the X′e vector (i.e. X11 ×e1 +X12 ×e2 +...+X1n ×en) to be zero, it must be the case that ei = 0.
The sample mean of the residuals is zero.	e= ∑e i/n = 0.
The regression hyperplane passes through the means of the observed values (X and y).	This follows from the fact that e = 0. Recall that e = y − Xβˆ. Dividing by the number of observations, we get e = y − xβˆ = 0. This implies that y = xβˆ. This shows that the regression hyperplane goes through the point of means of the data.
The predicted values of y are uncorrelated with the residuals.	ˆ′e = (Xβˆ)′e = b′X′e = 0
The mean of the predicted Y's for the sample will equal the mean of the observed Y's : y^-=y-	
The Gauss-Markov Theorem: Proof that βˆ is an unbiased estimator of β	βˆ = (X′X)$^{-1}$X′y=(X′X)$^{-1}$X′(Xβ + ε)
	β + (X′X)$^{-1}$X′ε
given (X′X)$^{-1}$X′X = I	E[βˆ] = E[β] + E[(X′X)$^{-1}$X′ε] = β + (X′X)$^{-1}$X′E[ε]
where E[X′ε]=0	E[βˆ]=β
Proof that βˆ is a linear estimator of β.	βˆ = β + (X′X)$^{-1}$X′ε; where (X′X)$^{-1}$X′= A
	βˆ = β + Aε => linear equation

Dummy Variables

Dummy/Binary Variables	= yes/no variables
	= take on the values 0 and 1 to identify the mutually exclusive classes of the explanatory variables.
	= leads to regression models where the parameters have very natural interpretations
Given: wage= $\beta_0 + \partial_0$ female + β_1 edu + u	
	To solve for ∂_0:
	∂_0=E(wage|female,edu)-E(wage|male,edu)
	where level of education is the same
Graphically ∂_0 =	an intercept shift
	male intercept= β_0
	female intercept= $\beta_0+\partial_0$
dummy variable trap=	when both dummy variables (male & female) are included; resulting in perfect collinearity
If a qualitative variable has m levels;	then $(m-1)$ dummy variables are required and each of them takes value 0 and 1.
Hypothesis test	
Test whether the two regression models are identical:	
	H0 : $\beta_2 = \beta_3$ 0
	H1 : $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$.
Acceptance of H0 indicates that only single model is necessary to explain the relationship.	
Test is two models differ with respect to intercepts only and they have same slopes	
	H0 : $\beta_3 = 0$
	H1: $\beta_3 \neq 0$.
Treating a **quantitative variable** as qualitative variable increases the complexity of the model.	
	The degrees of freedom for error are reduced.
	Can effect the inferences if data set is small

Inference

Normality Assumption:	zero mean and Variance
	Var(u)= σ^2
T-test:	$(\hat\beta_j - \beta_j)/se(\hat\beta_j) \sim t_{n-k-1}$ =t df
H0 : $\beta_j = 0$	used in testing hypotheses about a single population parameter as in .
Test statistic	$t_{\hat\beta_j} = (\hat\beta_j)/se(\hat\beta_j) \sim t_{n-k-1}$
	t = (estimate − hypothesised value)/ standard error
Alternative Hypothesis/one sided	
H1: $\beta_j > 0$	$t_{\hat\beta_j} > c$ [c @5%]
H1: $\beta_j < 0$	$t_{\hat\beta_j} <- c$ [c @5%]
Two sided	
H1: β_j =/= 0	$\|t_{\hat\beta_j}\| > c$ [c @2.5%]
If H0, rejected	x j is statistically significant, (significantly different from zero), @ the 5% level
if H0, not rejected	x j is statistically insignificant @the 5% level
P-value	smallest significant level at which the null hypotheses would be rejected
Confidence Interval	$\hat\beta_j \pm c \cdot se(\hat\beta_j)$
	where c is 97.5 percentile in a t_{n-k-1} distribution
CI given; @ 5% significant level	H0 : $\beta_j = a_j$ is rejected against H1: $\beta_j \not\leq a_j$; if aj is not in the 95% confidence interval
H0: $\beta_1 < \beta_2 \Leftrightarrow \beta_1 - \beta_2 < 0$	t= $(\hat\beta_1 - \hat\beta_2)$ /se$(\hat\beta_1 - \hat\beta_2)$
se$(\hat\beta_1 - \hat\beta_2) = \sqrt{Var(\hat\beta_1 - \hat\beta_2)}$	
	$Var(\hat\beta_1 - \hat\beta_2) = Var(\hat\beta_1) + Var(\hat\beta_2) - 2Cov(\hat\beta_1, \hat\beta_2)$
alternative to calculating se$(\hat\beta_1 - \hat\beta_2)$	Let $\theta = \hat\beta_1 - \hat\beta_2$; $\beta_1 = \theta + \hat\beta_2$
H0: θ=0, H1: θ<0	Substituting $\beta_1 = \theta + \hat\beta_2$ into the model we obtain

Inference (cont)

β0 +θ x1 +β2(x1 +x2)+β3 x3 +u	
F Test	F =[(SSRr-SSRur)/q] / [SSRur/(n-k-1)]
q	=number of restrictions
n-k-1= df ur	= df r- df ur
R^2 F stat	SSR= SST(1 - R^2)
	F= [(R^2ur-R^2r)/q] / [1-R^2ur)/(df ur)]
remember to not square the R value thats already been done	
Overall significance of the regression	
Testing joint exclusion	[R^2/R]/[(1-R^2)/(n-k-1)]

Data Scaling

Changes:	
if Xj is * by c	Its coefficient is / by c
If dependant variable is * by c	ALL OLS coefficients are * by c
neither t nor F statistics are affected	
Beta coefficients	obtained from an OLS regression after the dependant and independent variables have been transformed into z-scores