MFin Econometrics I

Session 6: Model specification errors and consequences, Heteroscedasticity robust estimation, Testing for Autocorrelation, Linearity and Normality

Thilo Klein

University of Cambridge
Judge Business School

Mis-specification in terms of variables

Consequences of mis-specification

		True Model	
		$Y = \beta_0 + \beta_1 X_1 + u$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
Fitted	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$		
	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$		

CAMBRIDGE Judge Business School

Mis-specification in terms of variables

Consequences of mis-specification

		True Model	
		$Y = \beta_0 + \beta_1 X_1 + u$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
Fitted	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$	Correct spec., no problems	
	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$		Correct specification, no problems

CAMBRIDGE Judge Business School

Mis-specification in terms of variables

Misspecification I: Omitting a relevant variable

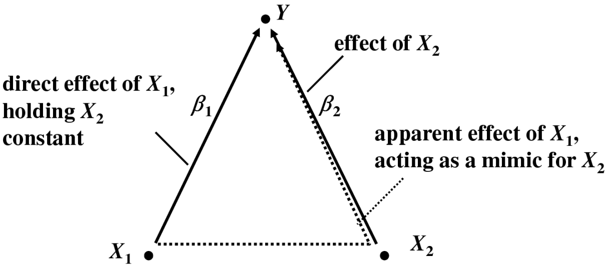
		True Model	
		$Y = \beta_0 + \beta_1 X_1 + u$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
Fitted	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$	Correct spec., no problems	Coefficients biased; Standard errors invalid
	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$		Correct specification, no problems

CAMBRIDGE Judge Business School

Mis-specification in terms of variables

Misspecification I: Omitting a relevant variable

$$\text{True model:} \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$
$$\text{Fitted model:} \quad \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$$



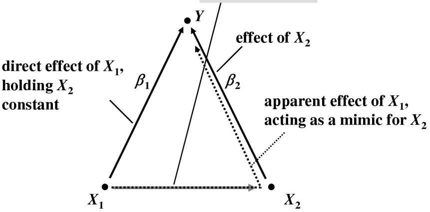
Mis-specification in terms of variables

Misspecification I: Omitting a relevant variable

$$\text{True model:} \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

$$\text{Fitted model:} \quad \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$$

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$$



Mis-specification in terms of variables

Misspecification II: Inclusion of an irrelevant variable

		True Model	
		$Y = \beta_0 + \beta_1 X_1 + u$	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
Fitted	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$	Correct spec., no problems	Coefficients biased; Standard errors invalid
Fitted	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$	Coefs inefficient; Std. errors are valid	Correct specification, no problems

CAMBRIDGE Judge Business School

Mis-specification in terms of variables

Misspecification II: Inclusion of an irrelevant variable

$$Y = \beta_0 + \beta_1 X_1 + u$$

- Let us say you include X_2 (which is irrelevant) as an explanatory variable

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

- What exactly have you estimated?
- Rewrite the true model adding X_2 as an explanatory variable, with a coefficient of 0

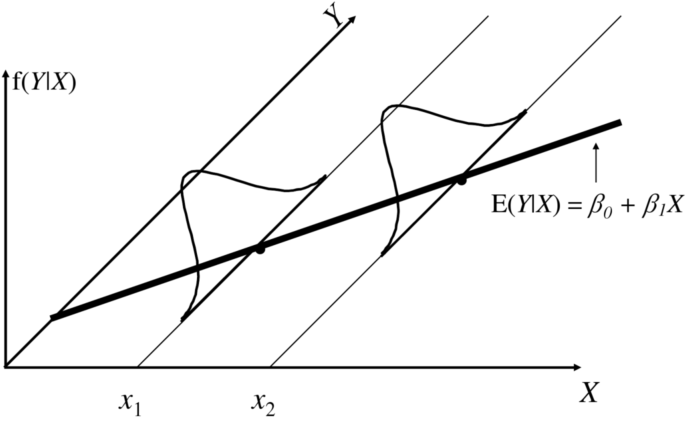
$$Y = \beta_0 + \beta_1 X_1 + 0X_2 + u$$

- Population variance of $\hat{\beta}_1 = \sigma^2_{\hat{\beta}_1} = \frac{\sigma^2_u}{nVar(X_1)} \times \frac{1}{1-r^2_{X_1,X_2}}$

CAMBRIDGE Judge Business School

Homoskedasticity example

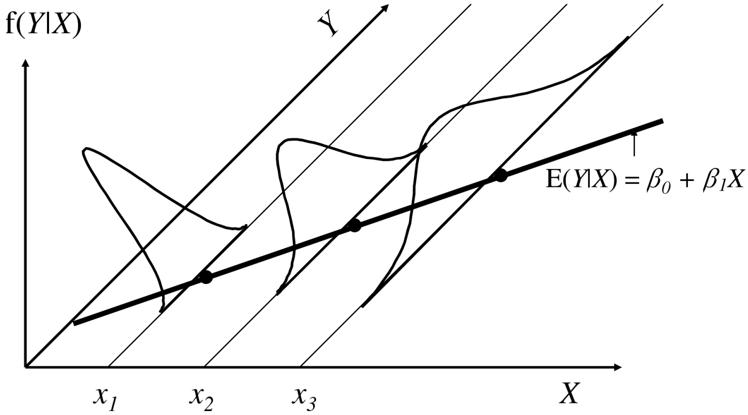
All observations have the same finite variance (homogeneity of variance)



CAMBRIDGE Judge Business School

Heteroskedasticity example

Implication is that while OLS estimators are unbiased, the std errors are larger than they need to be!



Tests for Heteroscedasticity

The Breusch-Pagan and White Tests

- Basic premise: if disturbances are homoscedastic, then squared errors are on average roughly constant.
- Regressors should NOT be able to predict squared errors, or their proxy, squared residuals.

Testing for Heteroscedasticity

Hypothesis

- Essentially, we want to test $H_0 : Var(u|X_1, X_2, ..., X_K) = \sigma^2$
- which is equivalent to $H_0 : E(u^2|X_1, X_2, ..., X_K) = \sigma^2$
- If we assume a possible linear relationship between u^2 and X_j, we can test:

$$H_0 : \delta_1 = \delta_2 = ... = \delta_k = 0$$

in the relationship

$$u^2 = \delta_0 + \delta_1 X_1 + ... + \delta_k X_k + v$$

CAMBRIDGE Judge Business School

Steps in the Breusch-Pagan Test

Breusch-Pagan Test

- Regress Y against explanators using OLS
- Compute the OLS residuals, $e_1, ..., e_n$
- Regress e_i^2 against a constant, all the explanators: $X_1, X_2, ..., X_k$

$$BP\ test\ stastic = nR^2$$

given the R^2 from this auxiliary regression.

- This is asymptoticaly distributed $\chi^2(k-1)$ under the null hypothesis of homoscedasticity.

Note: the above (Lagrange Multiplier) version of the test does *not* depend on normal distribution of disturbance terms

The White Test for Heteroscedasticity

An extension of the Breusch-Pagan test, the White test

- The White test allows **nonlinearities** by using squares and cross-products of all the X's in the auxiliary regression.
- A similar test statistic as before can be used to test whether all the x_j, x_j^2, and $x_j x_h$ are jointly significant.
- This can get to be unwieldy pretty quickly, burning through degrees of freedom very rapidly.
- Only appropriate for very large sample sizes.
- Failing this test (significant relationship between squares of residuals and squares and cross products of the explanatory variables) could also be an indication of misspecification of the functional form.
- Example

The mean and variance of the sampling distribution of $\hat{\beta}_1$ allowing heteroscedasticity

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

We have seen that:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$E(\hat{\beta}_1)$

$E(\hat{\beta}_1) - \beta_1 = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] = 0$; because $E(u_i|X_i = x) = 0$ by assumption, so $\hat{\beta}_1$ is an unbiased estimator of β_1.

What if the errors are in fact homoskedastic?

The formula for the variance of $\hat{\beta}_1$ and the OLS standard error simplifies: If $var(u_i|X_i = x) = \sigma_u^2$, then

$$var(\hat{\beta}_1) = \frac{var[(X_i - \mu_x)u_i]}{n(\sigma_x^2)^2} = \frac{E[(X_i - \mu_x)^2 u_i^2]}{n(\sigma_x^2)^2} = \frac{\sigma_u^2}{n\sigma_x^2}$$

OLS has lowest variance among estimators that are linear in Y.

Homoskedasticity only standard error formula

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

CAMBRIDGE Judge Business School

$SE(\hat{\beta}_1)$ if the errors are heteroscedastic

The expression for the variance of $\hat{\beta}_1$:

$$var(\hat{\beta}_1) = \frac{var[(X_i - \mu_x)u_i]}{n(\sigma_x^2)^2} = \frac{\sigma_v^2}{n\sigma_x^2},$$

where $v_i = (X_i - \mu_x)u_i$.
The estimator of the variance of $\hat{\beta}_1$ replaces the unknown population values of σ_v^2 and σ_x^2 by estimators constructed from the data:

$$\hat{\sigma}^2_{\beta_1} = \frac{1}{n} \times \frac{\hat{\sigma}_v^2}{\hat{\sigma}_x^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^n \hat{v}_i^2}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}$$

where $\hat{v}_i = (X_i - \bar{X})\hat{u}_i$.

CAMBRIDGE Judge Business School

What about the Sampling distribution of $\hat{\beta}_1$

Sampling distribution of $\hat{\beta}_1$

The exact sampling distribution of OLS estimators depends on the population distribution of (Y, X) - but when n is large we get some simple (and good) approximations:

1. Because $var(\hat{\beta}_1) \mapsto 1/n$ and $E(\hat{\beta}_1) = \beta_1$, $\hat{\beta}_1 \mapsto_p \beta_1$
2. When n is large, the sampling distribution of $\hat{\beta}_1$ is well approximated by a normal distribution (CLT)

Recall the CLT: suppose $\{v_i\}, i = 1, ..., n$ is i.i.d. with $E(v) = 0$ and $var(v) = \sigma_v^2$. Then, when n is large, $\frac{1}{n}\sum_{i=1}^{n} v_i$ is approximately distributed $N(0, \sigma_v^2/n)$.

So, Large-n approximation to the distribution of $\hat{\beta}_1$

Large-sample approximation

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma_v^2}{n\sigma_x^2}),$$

where $v_i = (X_i - \mu_x)u_i$.

- $\hat{\beta}_1$ is unbiased
- $var(\hat{\beta}_1)$ is inversely proportional to n

In summary, there are two formulas for standard errors for $\hat{\beta}_1$

homoscedasticity-only vs. "heteroscedasticity-robust"

- **Homoscedasticity-only** standard errors-these are valid only if the errors are homoskedastic.
- **Heteroscedasticity-robust** standard errors, valid whether or not the errors are heteroskedastic.
- The homoscedasticity-only formula for the standard error of $\hat{\beta}_1$ and the "heteroscedasticity-robust" formula differ-so in general, you get different standard errors using the different formulas.

Choice:

How to proceed?

- If the errors are either homoscedastic or heteroscedastic, you can use heteroscedastic-robust standard errors
- If the errors are heteroscedastic and you use the homoscedasticity-only formula for standard errors, your standard errors will be wrong (the homoscedasticity-only estimator of the variance of $\hat{\beta}_1$ is inconsistent if there is heteroscedasticity).
- The two formulas coincide (when n is large) in the special case of homoscedasticity

CAMBRIDGE Judge Business School

Robust estimation vs. Efficient estimation

Efficient estimation

- It is always possible to estimate robust standard errors for OLS estimates, BUT if we knew about the specific form of the heteroscedasticity, we could obtain more efficient estimates than OLS.
- The basic idea is to transform the model into one that has homoscedastic errors by weighting the squared residuals -therefore the name of the method, called weighted least squares.
- Methods: Weighted LS, Generalised LS, Feasible Generalised LS will be covered next term

Robustness vs. efficiency

a) Robust estimations:

not as dependent on assumptions but need large samples
Robust standard errors only have asymptotic (large sample) justification -with small sample sizes, inferences will not be correct

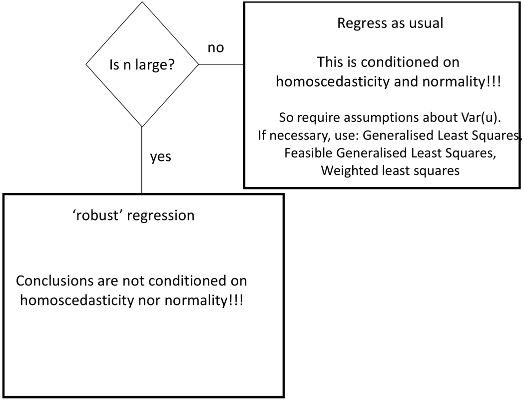
b) Efficient estimations:

need to incorporate the explicit specification (if known) of the disturbances into the model

- If the specification is correct, then, b) is more efficient.
- If the sample is large, then, a) is satisfactory

Decision tree

Robust or as usual?



Autocorrelation (Serial correlation)

IMPLICATION: Knowledge of one residual, helps to predict other residual(s)

Usual cause: Misspecification

- Omitted variable - a serially correlated explanatory variable is omitted
- Incorrect functional form

Consequence:

- OLS estimators unbiased and consistent
- OLS estimators not efficient.
- Standard errors are wrong. Generally under-estimated. "t-statistics" tend to be higher.

CAMBRIDGE Judge Business School

Autocorrelation: a common pattern AR(1)

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

First-order serial correlation (or auto correlation) : AR(1)

$$u + i = \rho u_{i-1} + \epsilon_i \quad -1 < \rho < 1$$

with ε_i, white noise (the ε_i are independent and all have the same variance and mean 0). Note: the autocorrelation process may be more general, over k lags.

Preliminary diagnosis:

- OLS Time series graph of e_i, $t = 1, ..., n$.
- Scatterplot of e_i on e_{i-1}. If AR(1) model $u_i = \rho u_{i-1} + e_i$ holds, then we expect the scatterplot to be concentrated along a straight line through 0.

CAMBRIDGE Judge Business School

Autocorrelation

A test:

- If $\rho = 0$, then $u_i = \varepsilon_i$ and in that case the random errors u_i satisfy the i.i.d. assumption (no serial correlation).
- Hence a test for serial correlation is the test: $H_0 : \rho = 0$.
- First step is to find estimator for ρ. If we replace u_i in the AR(1) model for the disturbance by e_i and estimate ρ by OLS, we obtain:

$$\rho = \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}$$

- This is also the first-order autocorrelation coefficient of the time series e_i , $i = 1, ..., n$. The obvious thing to do is to use $\hat{\rho}$ to test whether $\rho = 0$. Instead of $\hat{\rho}$, a related quantity is used, the Durbin-Watson statistic d.

CAMBRIDGE Judge Business School

Durbin-Watson test for AR(1) autocorrelation

$$d = \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}$$

Test statistic

- It can be shown that $d = 2(1 - \hat{\rho})$
- Hence if $\hat{\rho}$ is close to 0 (no autocorrelation) then d is close to 2.
- If $\hat{\rho}$ is close to 1, then d is close to 0 and if $\hat{\rho}$ is close to -1, then d is close to 4.
- In large samples $d \mapsto 2 - 2\rho$
- No autocorrelation $d \mapsto 2$

CAMBRIDGE Judge Business School

Durbin-Watson test for AR(1) autocorrelation

Durbin-Watson test for AR(1) autocorrelation

- Critical values for the test of the null hypothesis of no autocorrelation depends on the specific values taken by the explanatory variables
- But lower and upper bounds for the critical values that do not depend on the X have been calculated by Durbin and Watson.
- Assumptions for the test: Constant term included; Normal disturbances, Lagged values of the dependent variable is not included as an explanator

Breusch-Godfrey test for Autocorrelation

Durbin-Watson test for AR(1) autocorrelation

- An alternative to DW test is another (Lagrange Multiplier - LM) test which is also based on OLS residuals, e_i.
- The first step is an auxiliary linear regression with dependent variable e_i and independent variables $X_1, ..., X_K$, and e_{i-1}. Compute the R^2 of this regression.
- The test statistic is $LM = (n-1)R^2$
  Note: we use n -1 observations in this auxiliary regression.
- If H_0:$\rho = 0$ is true than LM has a $\chi^2(1)$ distribution: 1d.f.
- We reject the null if LM > c, the critical value found from the χ^2 distribution.

Note: this is a test for the AR(1) form of autocorrelation. Will generalise to AR(p), next term.

CAMBRIDGE Judge Business School

Linearity

In a bivariate relationship:

- scatter plot between the response variable and the predictor to see if nonlinearity is present.

In a multivariate relationship:

- Plot the residuals against each of the predictor variables in the regression model. If there is a clear nonlinear pattern, there is a problem of nonlinearity.
- We should see a random scatter of points, for each plots.
- Can estimate a **locally weighted regression** of residual on each explanatory variable to diagnose.
- Often a log transformation of positively skewed explanatory variables can help.

Another example.

CAMBRIDGE Judge Business School

Nonlinearity as an error in Model specification

Regression specification error test (RESET) for non-linearity

The test is based on creating new variables (based on the predictors) and refitting the model using those new variables to see if any of them would be significant.

- Procedure:
  - Regress Y on $X_1, X_2, ..., X_k$, obtain $\hat{Y}$
  - Calculate powers of the fitted values $\hat{Y}^2, \hat{Y}^3, \hat{Y}^4$
  - Refit the model with original regressors and these new powers of predicted values from the original regression, and powers of original regressors.
  - Under the **null that there is no functional form mis-specification**, that coefficients on these new variables should be zero. RESET is a test of this joint hypothesis.
- Logic: Polynomials in $\hat{Y}$ and X_j can approximate a variety of non-linear relationships between Y and $X_1, X_2, ..., X_k$

More general model specification errors

More general model specification errors

- A model specification error can occur when one or more relevant variables are omitted from the model or one or more irrelevant variables are included in the model.
- There is no direct test for this type of error.

Normality

Normality

- Normality is not required in order to obtain unbiased estimates of the regression coefficients.
    - But normality of the residuals necessary for some of the other tests – for example the Breusch Pagan test of heteroscedasticity, Durbin Watson test of autocorrelation, etc.
    - Note: there is no requirement that the predictor variables be normally distributed. But regression is more effective if the predictor variables have a roughly symmetric distribution with a single mode and no outliers.
- After regression estimation, can obtain residuals.
- And then can use various tests for normality of the residuals.

CAMBRIDGE Judge Business School

Normality

Normality

- Estimate the regression model
- Examine the residuals.
- Examine the histogram plot against the normal density overlaid on the plot.
- Kernel density: a histogram with narrow bins and moving average. Kernel density is the smoothed out contribution of each observed data point over a local neighbourhood of that data point.

CAMBRIDGE Judge Business School

Normality

Visualisation - pp-plot

- The **pp-plot** graphs a standardized normal probability plot.
- In a normal probability plot, the data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line.
- Departures from this straight line indicate departures from normality.
  1. The data are arranged from smallest to largest.
  2. The percentile of each data value is determined.
  3. the z-score of each data value is calculated.
  4. z-scores are plotted against the percentiles of data values

PP-plot is sensitive to non-normality in the middle range of data.

CAMBRIDGE Judge Business School

Normality

Visualisation - qq-plot

- **qq-plots** plot the quantiles of a variable against the quantiles of a normal distribution.
- qq plot is sensitive to non-normality near the tails.
- The results from p-p plots and q-q plots show indications of non-normality,
- You could examine the results from p-p plots and q-q plots after different specifications that make sense (linear regression, double log regression, semi-log regression etc)
- Can form a judgement about it is possible to accept that the residuals are close to a normal distribution.

Jarque-Bera test of normality

Jarque-Bera test of normality

- A standard normality test. There are others too.
- The test is based on properties on skewness and kurtosis of the normal distribution (Skew=0 and Kurtosis =3)
- Null Hypothesis: the residuals are normally distributed.
- Deviation from normality measured by:

$$JB = n(\frac{1}{6}S^2 + \frac{1}{24}(K-3)^2) \mapsto_d \chi^2_2$$

- JB statistic is distributed χ^2_2 under the Null hypothesis
- Reject normality if the p-value is below your chosen test size

CAMBRIDGE Judge Business School

Repairing Normality

Repairing Normality

- What is the pattern in the plot of residuals?
- Check alternative (non-linear) specifications that are appropriate.
- Deviations from normality could be due to outliers.
  - Find the reasons for outliers.
  - Data error? Correct the entry.
  - If not data error, and there is a **valid reason** for that observation,
  - then could use a dummy variable for that observation.