Financial econometrics

Chapter 4, Classical linear regression model assumptions and diagnostics

Overview

1. Introduction
2. Residuals: zero-mean, heteroskedasticity and autocorrelation
   - Residuals' mean
   - Heteroscedasticity
   - Goldfeld-Quandt test
   - White's test
   - Dealing with heteroscedasticity
   - Autocorrelation
   - Durbin-Watson Test
   - Breusch-Godfrey test
   - Remedies for Autocorrelation
   - Dynamic models
3. Multicollinearity and wrong functional form
   - Multicollinearity
   - Wrong functional form
4. Normality
5. Omitted variables
   - Omitted variable
   - Irrelevant variable
6. Parameter stability
   - Parameter stability
   - Chow test
   - Predictive failure test
7. Error-in-variable models
   - Measurement error
   - Measurement error in CAPM
8. References

Section 1, Introduction

Violation of the Assumptions of the CLRM

- Recall what we assumed for the Classical linear regression model (CLRM) and diagnostics disturbance terms:
  1. $\mathrm{E}(u_t) = 0$
  2. $\mathrm{var}(u_t) = \sigma^2 < \infty$
  3. $\mathrm{cov}(u_i, u_j) = 0$
  4. The X matrix is non-stochastic or fixed in repeated samples $\mathrm{cov}(u_t, x_t) = 0$
  5. $u_t \sim \mathrm{N}(0,\ \sigma^2)$

Investigating Violations of the Assumptions of the CLRM

- We will now study these assumptions further, and in particular look at:
  - How we test for violations
  - Causes
  - Consequences

    in general we could encounter any combination of 3 problems:
    - the coefficient estimates are wrong
    - the associated standard errors are wrong
    - the distribution that we assumed for the test statistics will be inappropriate
  - Solutions
  - the assumptions are no longer violated
  - we work around the problem so that we use alternative techniques which are still valid

Statistical Distributions for Diagnostic Tests

- Often, an *F*- and a χ^2- version of the test are available.
- The *F*-test version involves estimating a restricted and an unrestricted version of a test regression and comparing the *RSS*.
- The χ^2- version is sometimes called an "LM" test, and only has one degree of freedom parameter: the number of restrictions being tested, *m*.
- Asymptotically, the 2 tests are equivalent since the χ^2 is a special case of the *F*-distribution:

$$\frac{\chi^2(m)}{m} \rightarrow F(m, T-k) \quad \text{as} \quad (T-k) \rightarrow \infty$$

- For small samples, the *F*-version is preferable.

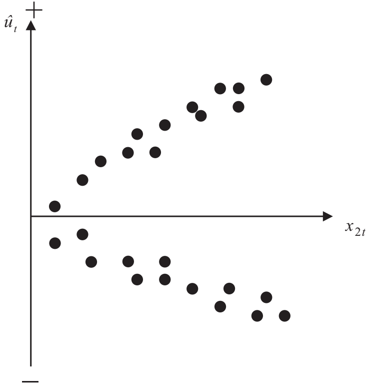
Section 2, Residuals: zero-mean, heteroskedasticity and autocorrelation

Assumption 1: $E(u_t) = 0$

- Assumption that the mean of the disturbances is zero.
- For all diagnostic tests, we cannot observe the disturbances and so perform the tests of the residuals.
- The mean of the residuals will always be zero provided that there is a constant term in the regression.

Assumption 2: $\text{var}(u_t) = \sigma^2 < \infty$

- We have so far assumed that the variance of the errors is constant, σ^2 - this is known as homoscedasticity. If the errors do not have a constant variance, we say that they are heteroscedastic e.g. say we estimate a regression and calculate the residuals, $\hat{u}_t$.



Detection of Heteroscedasticity

- Graphical methods
- Formal tests: There are many of them: we will discuss Goldfeld-Quandt (Goldfeld and Quandt, 1965) test and White's test (White, 1980)

[1] Stephen M Goldfeld and Richard E Quandt. 'Some tests for homoscedasticity'. In: *Journal of the American statistical Association* 60.310 (1965), pp. 539–547

[2] Halbert White. 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity'. In: *Econometrica: journal of the Econometric Society* (1980), pp. 817–838

The GQ test I

The Goldfeld-Quandt (GQ) test is carried out as follows.

1. Split the total sample of length T into two sub-samples of length T_1 and T_2. The regression model is estimated on each sub-sample and the two residual variances are calculated.

2. The null hypothesis is that the variances of the disturbances are equal, $H_0 : \sigma_1^2 = \sigma_2^2$

3. The test statistic, denoted GQ, is simply the ratio of the two residual variances where the larger of the two variances must be placed in the numerator.

$$GQ = \frac{s_2^2}{s_1^2} = \frac{RSS_2/(T_2 - k)}{RSS_1(T_1 - k)}$$

4. The test statistic is distributed as an $\mathrm{F}(T_1 - k,\ T_2 - k)$ under the null of homoscedasticity.

The GQ test II

5. A problem with the test is that the choice of where to split the sample is that usually arbitrary and may crucially affect the outcome of the test.

Detection of Heteroscedasticity using White's Test I

- White's general test for heteroscedasticity is one of the best approaches because it makes few assumptions about the form of heteroscedasticity.
- The test is carried out as follows:
  1. Assume that the regression we carried out is as follows

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

     And we want to test $\text{Var}(u_t) = \sigma^2$. We estimate the model, obtaining the residuals, $\hat{u}_t$.

  2. Then run the auxiliary regression

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

Detection of Heteroscedasticity using White's Test II

3. Obtain the R^2 from the auxiliary regression and multiply it by the number of observations, T. It can be shown that

$$T \times R^2 \sim \chi^2(m)$$

   where m is the number of regressors in the auxiliary regression excluding the constant term.

4. If the χ^2 test statistic from step 3 is greater than the corresponding value from the statistical table then reject the null hypothesis that the disturbances are homoscedastic.

Consequences of Using OLS in the Presence of Heteroscedasticity

- OLS estimation still gives unbiased coefficient estimates, but they are no longer BLUE.
- This implies that if we still use OLS in the presence of heteroscedasticity, our standard errors could be inappropriate and hence any inferences we make could be misleading.
- Whether the standard errors calculated using the usual formulae are too big or too small will depend upon the form of the heteroscedasticity.

How Do we Deal with Heteroscedasticity? I

- If the form (i.e. the cause) of the heteroscedasticity is known, then we can use an estimation method which takes this into account (called generalised least squares, GLS).
- A simple illustration of GLS is as follows: Suppose that the error variance is related to another variable z_t by

$$\mathrm{var}(u_t) = \sigma^2 z_t^2$$

- To remove the heteroscedasticity, divide the regression equation by z_t

$$\frac{y_t}{z_t} = \beta_1 \frac{1}{z_t} + \beta_2 \frac{x_{2t}}{z_t} + \beta_3 \frac{x_{3t}}{z_t} + v_t$$

where $v_t = \dfrac{u_t}{z_t}$ is an error term.

- Now $\mathrm{var}(u_t) = \sigma^2 z_t^2$, $\text{var}(v_t) = \text{var}\left(\dfrac{u_t}{z_t}\right) = \dfrac{\text{var}(u_t)}{z_t^2} = \dfrac{\sigma^2 z_t^2}{z_t^2} = \sigma^2$ for known z_t.

Other Approaches to Dealing with Heteroscedasticity

- So the disturbances from the new regression equation will be homoscedastic.

- Other solutions include:
  1. Transforming the variables into logs or reducing by some other measure of “size”.
  2. Use White’s heteroscedasticity consistent standard error estimates.

  The effect of using White’s correction is that in general the standard errors for the slope coefficients are increased relative to the usual OLS standard errors.

  This makes us more “conservative” in hypothesis testing, so that we would need more evidence against the null hypothesis before we would reject it.

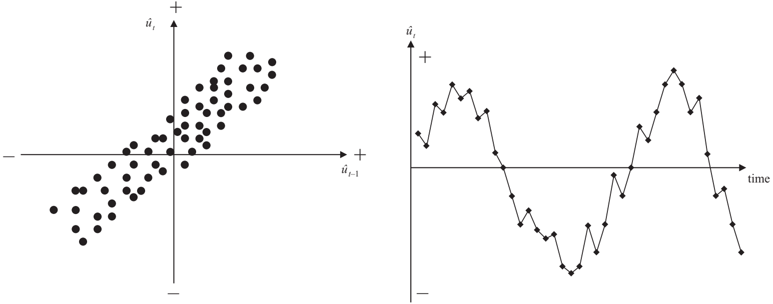
Background – The Concept of a Lagged Value

t	y_t	y_{t-1}	Δy_t
2006M09	0.8	–	–
2006M10	1.3	0.8	$(1.3 - 0.8) = 0.5$
2006M11	-0.9	1.3	$(-0.9 - 1.3) = -2.2$
2006M12	0.2	-0.9	$(0.2 - -0.9) = 1.1$
2007M01	-1.7	0.2	$(-1.7 -0.2) = -1.9$
2007M02	2.3	-1.7	$(2.3 - -1.7) = 4.0$
2007M03	0.1	2.3	$(0.1 - 2.3) = -2.2$
2007M04	0.0	0.1	$(0.0 - 0.1) = -0.1$
.	.	.	.
.	.	.	.
.	.	.	.

Autocorrelation

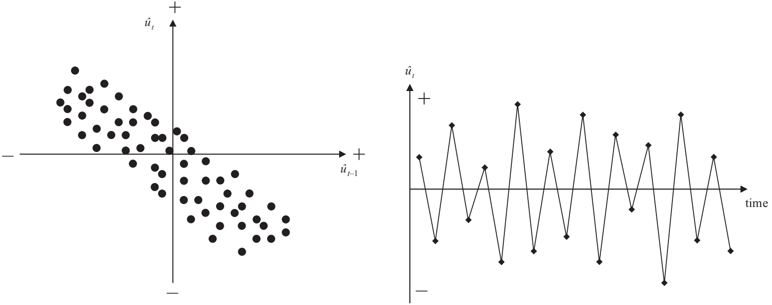
- We assumed of the CLRM's errors that Cov $(u_i, u_j) = 0$ for $i \neq j$, This is essentially the same as saying there is no pattern in the errors.
- Obviously we never have the actual u's, so we use their sample counterpart, the residuals (the $\hat{u}_t$'s).
- If there are patterns in the residuals from a model, we say that they are autocorrelated.
- Some stereotypical patterns we may find in the residuals are given on the next 3 slides.

Positive Autocorrelation



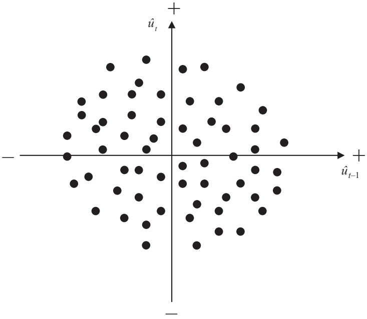
Positive Autocorrelation is indicated by a cyclical residual plot over time.

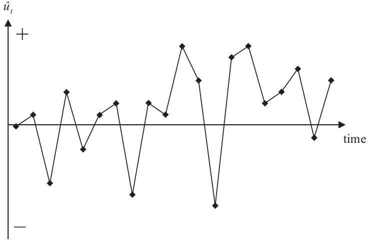
Negative Autocorrelation



Negative autocorrelation is indicated by an alternating pattern where the residuals cross the time axis more frequently than if they were distributed randomly

No pattern in residuals – No autocorrelation





No pattern in residuals at all: this is what we would like to see

The Durbin-Watson Test

[1] James Durbin and Geoffrey S Watson. 'Testing for serial correlation in least squares regression. I'. In: *Breakthroughs in Statistics*. Springer, 1992, pp. 237–259

- The Durbin-Watson (DW) is a test (Durbin and Watson, 1992) for first order autocorrelation - i.e. it assumes that the relationship is between an error and the previous one

$$u_t = \rho u_{t-1} + v_t \tag{1}$$

where $v_t \sim N(0, \sigma_v^2)$.

The Durbin-Watson Test (cont'd)

- The DW test statistic actually tests

$$\mathrm{H}_0 : \rho = 0 \quad \text{and} \quad \mathrm{H}_1 : \rho \neq 0$$

- The test statistic is calculated by

$$DW = \sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2 / \sum_{t=2}^{T} \hat{u}_t^2$$

The Durbin-Watson Test: Critical Values

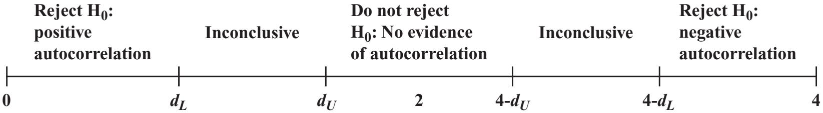
- We can also write
$$DW \approx 2(1 - \hat{\rho}) \tag{2}$$
where $\hat{\rho}$ is the estimated correlation coefficient. Since $\hat{\rho}$ is a correlation, it implies that $-1 \leq \hat{\rho} \leq 1$.

- Rearranging for DW from (2) would give $0 \leq DW \leq 4$.

- If $\hat{\rho} = 0$, DW=2. So roughly speaking, do not reject the null hypothesis if DW is near 2 → i.e. there is little evidence of autocorrelation

- Unfortunately, DW has 2 critical values, an upper critical value (d_U) and a lower critical value (d_L), and there is also an intermediate region where we can neither reject nor not reject H_0.

The Durbin-Watson Test: Interpreting the Results



Reject H_0: positive autocorrelation

Inconclusive

Do not reject H_0: No evidence of autocorrelation

Inconclusive

Reject H_0: negative autocorrelation

0 | d_L | d_U | 2 | $4\text{-}d_U$ | $4\text{-}d_L$ | 4

Conditions which Must be Fulfilled for DW to be a Valid Test

1. Constant term in regression
2. Regressors are non-stochastic
3. No lags of dependent variable

Another Test for Autocorrelation: The Breusch-Godfrey Test I

- It is a more general test for r^{th} order autocorrelation:

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \cdots + \rho_r u_{t-r} + v_t, \\ & v_t \sim N\left(0, \sigma_v^2\right) \end{aligned}$$

- The null and alternative hypotheses are:

$$\begin{aligned} H_0 &: \rho_1 = 0 \ \text{ and } \ \rho_2 = 0 \ \text{ and } \ldots \text{and } \rho_r = 0 \\ H_1 &: \rho_1 \neq 0 \ \text{ or } \ \rho_2 \neq 0 \ \text{ or } \ldots \text{or } \rho_r \neq 0 \end{aligned}$$

- The test is carried out as follows:
  1. Estimate the linear regression using OLS and obtain the residuals, $\hat{u}_t$.

Another Test for Autocorrelation: The Breusch-Godfrey Test II

2. Regress $\hat{u}_t$ on all of the regressors from stage 1 (the xs) plus $\hat{u}_{t-1}$, $\hat{u}_{t-2}, \ldots, \hat{u}_{t-r}$;
   Obtain R^2 from this regression.
3. It can be shown that

$$(T-r)R^2 \sim \chi^2_r$$

- If the test statistic exceeds the critical value from the statistical tables, reject the null hypothesis of no autocorrelation.

Consequences of Ignoring Autocorrelation if it is Present

- The coefficient estimates derived using OLS are still unbiased, but they are inefficient, i.e. they are not BLUE, even in large sample sizes.

- Thus, if the standard error estimates are inappropriate, there exists the possibility that we could make the wrong inferences.

- R^2 is likely to be inflated relative to its "correct" value for positively correlated residuals.

“Remedies” for Autocorrelation

- If the form of the autocorrelation is known, we could use a GLS procedure – i.e. an approach that allows for autocorrelated residuals e.g., Cochrane-Orcutt estimation (Cochrane and Orcutt, 1949).
- But such procedures that “correct” for autocorrelation require assumptions about the form of the autocorrelation.
- If these assumptions are invalid, the cure would be more dangerous than the disease! - see Hendry and Mizon, 1978.
- However, it is unlikely to be the case that the form of the autocorrelation is known, and a more “modern” view is that residual autocorrelation presents an opportunity to modify the regression.

Dynamic Models

- All of the models we have considered so far have been static, e.g.

$$y_t = \beta_1 + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t$$

- But we can easily extend this analysis to the case where the current value of y t depends on previous values of y or one of the x's, e.g.

$$\begin{aligned} y_t &= \beta_1 + \beta_2 \, x_{2t} + \cdots + \beta_k \, x_{kt} + \gamma_1 y_{t-1} + \gamma_2 x_{2t-1} \\ &\quad + \cdots + \gamma_k x_{kt-1} + u_t \end{aligned}$$

- We could extend the model even further by adding extra lags, e.g. x_{2t-2}, y_{t-3}.

Why Might we Want/Need To Include Lags in a Regression?

- Inertia of the dependent variable
- Over-reactions
- Measuring time series as overlapping moving averages
- However, other problems with the regression could cause the null hypothesis of no autocorrelation to be rejected:
  - Omission of relevant variables, which are themselves autocorrelated.
  - If we have committed a “misspecification” error by using an inappropriate functional form.
  - Autocorrelation resulting from unparameterised seasonality.

Models in First Difference Form

- Another way to sometimes deal with the problem of autocorrelation is to switch to a model in first differences.
- Denote the first difference of y_t, i.e. $y_t - y_{t-1}$ as Δy_t; similarly for the x-variables, $\Delta x_{2t} = x_{2t} - x_{2t-1}$ etc.
- The model would now be

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \cdots \beta_k \Delta x_{kt} + u_t$$

- Sometimes the change in y is purported to depend on previous values of y or x_t as well as changes in x:

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \beta_3 \Delta x_{2t-1} + \beta_4 y_{t-1} + u_t$$

Problems with Adding Lagged Regressors to "Cure" Autocorrelation

- Inclusion of lagged values of the dependent variable violates the assumption that the RHS variables are non-stochastic.
- What does an equation with a large number of lags actually mean?
- Note that if there is still autocorrelation in the residuals of a model including lags, then the OLS estimators will not even be consistent.

Section 3, Multicollinearity and wrong functional form

Multicollinearity

- This problem occurs when the explanatory variables are very highly correlated with each other.

- Perfect multicollinearity

  Cannot estimate all the coefficients

  - e.g. suppose $x_3 = 2x_2$
    and the model is $y_t = \beta_1 + \beta_2 x_{2t} + \beta x_{3t} + \beta_4 x_{4t} + u_t$

- Problems if Near Multicollinearity is Present but Ignored
  - R^2 will be high but the individual coefficients will have high standard errors.
  - The regression becomes very sensitive to small changes in the specification.
  - Thus confidence intervals for the parameters will be very wide, and significance tests might therefore give inappropriate conclusions.

Measuring Multicollinearity

- The easiest way to measure the extent of multicollinearity is simply to look at the matrix of correlations between the individual variables. e.g.

corr	x_2	x_3	x_4
x_2	–	0.2	$\underline{0.8}$
x_3	0.2	–	0.3
x_4	$\underline{0.8}$	0.3	–

- But another problem: if 3 or more variables are linear
  - e.g. $x_{2t} + x_{3t} = x_{4t}$

- Note that high correlation between y and one of the x's is not muticollinearity.

Solutions to the Problem of Multicollinearity

- “Traditional” approaches, such as ridge regression or principal components. But these usually bring more problems than they solve.
- Some econometricians argue that if the model is otherwise OK, just ignore it
- The easiest ways to “cure” the problems are
  - drop one of the collinear variables
  - transform the highly correlated variables into a ratio
  - go out and collect more data e.g.
    - a longer run of data
    - switch to a higher frequency

Adopting the Wrong Functional Form

- We have previously assumed that the appropriate functional form is linear.
- This may not always be true.
- We can formally test this using Ramsey's RESET test (Ramsey, 1969), which is a general test for mis-specification of functional form.
- Essentially the method works by adding higher order terms of the fitted values (e.g. $\hat{y}_t^2, \hat{y}_t^3$, etc.) into an auxiliary regression:

  Regress $\hat{u}_t$ on powers of the fitted values:

  $$\hat{u}_t = \beta_0 + \beta_1 \hat{y}_t^2 + \beta_2 \hat{y}_t^3 + \cdots + \beta_{p-1} \hat{y}_t^p + v_t$$

  Obtain R^2 from this regression. The test statistic is given by TR^2 and is distributed as a $\chi^2(p-1)$.
- So if the value of the test statistic is greater than a $\chi^2(p-1)$ then reject the null hypothesis that the functional form was correct.

But what do we do if this is the case?

- The RESET test gives us no guide as to what a better specification might be.
- One possible cause of rejection of the test is if the true model is

  $$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t}^2 + \beta_4 x_{2t}^3 + u_t$$

  In this case the remedy is obvious.
- Another possibility is to transform the data into logarithms. This will linearise many previously multiplicative models into additive ones:

  $$y_t = A x_t^{\beta} e^{u_t} \Leftrightarrow \ln(y_t) = \alpha + \beta \ln(x_t) + u_t$$

Section 4,
Normality

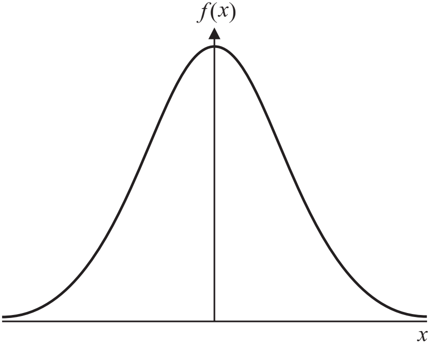
Testing the Normality Assumption

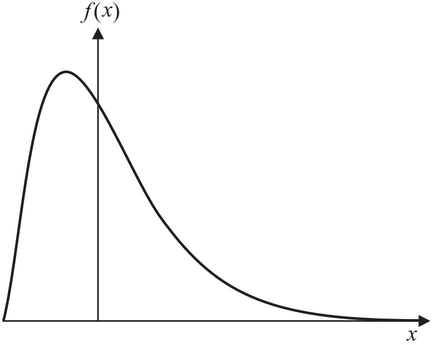
- Why did we need to assume normality for hypothesis testing?

Testing for Departures from Normality

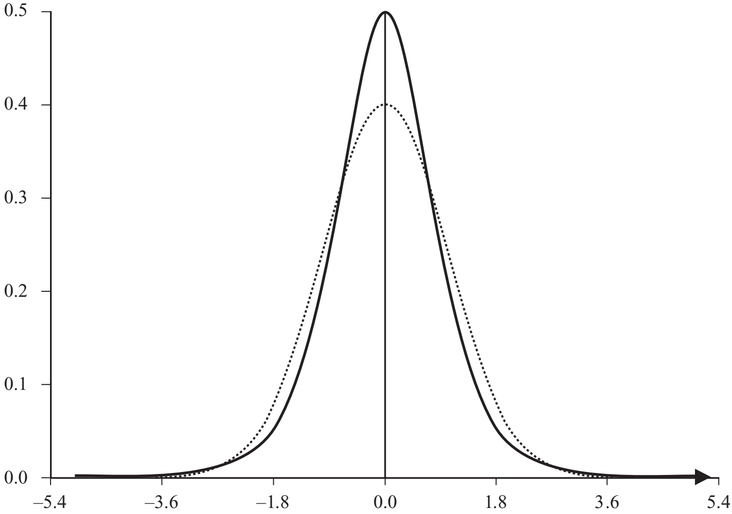
- *The Bera Jarque normality test* (Jarque and Bera, 1980)
- A normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3.
- The kurtosis of the normal distribution is 3 so its excess kurtosis (b_2-3) is zero.
- Skewness and kurtosis are the (standardised) third and fourth moments of a distribution.

Normal versus Skewed Distributions





Leptokurtic versus Normal Distribution



Testing for Normality

- Bera and Jarque formalise this by testing the residuals for normality by testing whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero.
- It can be proved that the coefficients of skewness and kurtosis can be expressed respectively as:

$$b_1 = \frac{E[u^3]}{(\sigma^2)^{3/2}} \quad \text{and} \quad b_2 = \frac{E[u^4]}{(\sigma^2)^2}$$

- The Bera Jarque test statistic is given by

$$W = T\left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24}\right] \sim \chi^2$$

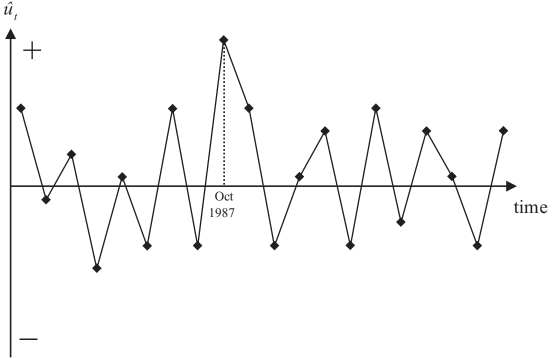
- We estimate b_1 and b_2 using the residuals from the OLS regression, .

What do we do if we find evidence of Non-Normality?

- It is not obvious what we should do!
- Could use a method which does not assume normality, but difficult and what are its properties?
- Often the case that one or two very extreme residuals causes us to reject the normality assumption.
- An alternative is to use dummy variables.

  e.g. say we estimate a monthly model of asset returns from 1980-1990, and we plot the residuals, and find a particularly large outlier for October 1987:

What do we do if we find evidence of Non-Normality? (cont'd)



- Create a new variable:
  $D87M10_t = 1$ during October 1987 and zero otherwise.
  This effectively knocks out that observation. But we need a theoretical reason for adding dummy variables.

Section 5,
Omitted variables

Omission of an Important Variable or Inclusion of an Irrelevant Variable

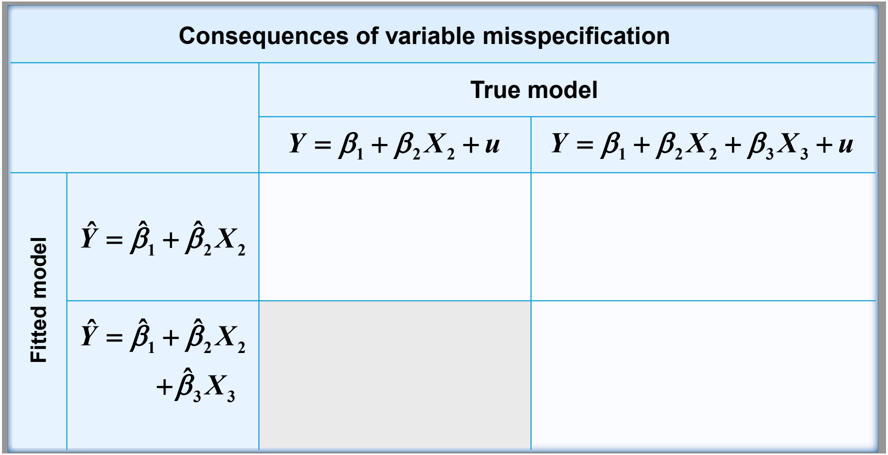
Omission of an Important Variable

- Consequence: The estimated coefficients on all the other variables will be biased and inconsistent unless the excluded variable is uncorrelated with all the included variables.
- Even if this condition is satisfied, the estimate of the coefficient on the constant term will be biased.
- The standard errors will also be biased.

Inclusion of an Irrelevant Variable

- Coefficient estimates will still be consistent and unbiased, but the estimators will be inefficient.

Omission of an Important Variable or Inclusion of an Irrelevant Variable



Consequences of variable misspecification

		True model	
		$Y = \beta_1 + \beta_2 X_2 + u$	$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$
Fitted model	$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2$		
	$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$		

Figure: Problem statement.

Omitted Variable

Omitted-variable bias (OVB) occurs when a statistical model leaves out one or more relevant variables. The bias results in the model attributing the effect of the missing variables to those that were included.

OVB is the bias that appears in the estimates of parameters in a regression analysis, when the assumed specification is incorrect in that it omits an independent variable which is:

1. a determinant of the dependent variable
2. correlated with one or more of the included independent variables

The true model is:

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

Omitted Variable (cont.)

But we estimate:

$$y = \beta_1 + \beta_2 X_2 + u$$

Two conditions must hold true for omitted-variable bias to exist in linear regression:

- the omitted variable must be a determinant of the dependent variable (i.e., its true regression coefficient must not be zero, $\beta_3 \neq 0$);
- the omitted variable must be correlated with an independent variable specified in the regression (i.e., $cov(X_2, X_3) \neq 0$).

Omitted Variable (cont.)

$$\hat{\beta}_2 = \frac{\sum \left(X_{2i} - \bar{X}_2\right)\left(Y_i - \bar{Y}\right)}{\sum \left(X_{2i} - \bar{X}_2\right)^2}$$

Although Y really depends on X_3 as well as X_2, we make a mistake and regress Y on X_2 only. The slope coefficient is therefore as shown above.

Where is the bias?

Omitted Variable (cont.)

$$
\begin{aligned}
\hat{\beta}_2 &= \frac{\sum\left(X_{2i}-\bar{X}_2\right)\left(Y_i-\bar{Y}\right)}{\sum\left(X_{2i}-\bar{X}_2\right)^2} \\
&= \frac{\sum\left(\beta_2\left(X_{2i}-\bar{X}_2\right)^2+\beta_3\left(X_{2i}-\bar{X}_2\right)\left(X_{3i}-\bar{X}_3\right)+\left(X_{2i}-\bar{X}_2\right)\left(u_i-\bar{u}\right)\right)}{\sum\left(X_{2i}-\bar{X}_2\right)^2} \\
&= \beta_2+\beta_3\frac{\sum\left(X_{2i}-\bar{X}_2\right)\left(X_{3i}-\bar{X}_3\right)}{\sum\left(X_{2i}-\bar{X}_2\right)^2}+\frac{\sum\left(X_{2i}-\bar{X}_2\right)\left(u_i-\bar{u}\right)}{\sum\left(X_{2i}-\bar{X}_2\right)^2}
\end{aligned}
$$

Omitted Variable (cont.)

From which:

$$\mathbb{E}\left[\hat{\beta}_2\right] = \beta_2 + \beta_3 \frac{\sum\left(X_{2i} - \bar{X}_2\right)\left(X_{3i} - \bar{X}_3\right)}{\sum\left(X_{2i} - \bar{X}_2\right)^2}$$

- Thus we have shown that the expected value of $\hat{\beta}_2$ is equal to the true value plus a bias term.
  Note: the definition of a bias is the difference between the expected value of an estimator and the true value of the parameter being estimated.
- As a consequence of the misspecification, the standard errors, *t*-tests and *F*-test are invalid.

Inclusion of an Irrelevant Variable

The true model is:

$$y = \beta_1 + \beta_2 X_2 + u,$$

but we estimate:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3,$$

for some X_3 that is irrelevant for explaining y.

Where's the problem?

Inclusion of an Irrelevant Variable (cont.)

Bias:

Rewrite the true model adding X_3 as an explanatory variable, with a coefficient of 0. Now the true model and the fitted model coincide. Hence $\hat{\beta}_2$ will be an unbiased estimator of β_2 and $\hat{\beta}_3$ will be an unbiased estimator of 0.

$$y = \beta_1 + \beta_2 X_2 + 0 X_3 + u$$
$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

Inclusion of an Irrelevant Variable (cont.)

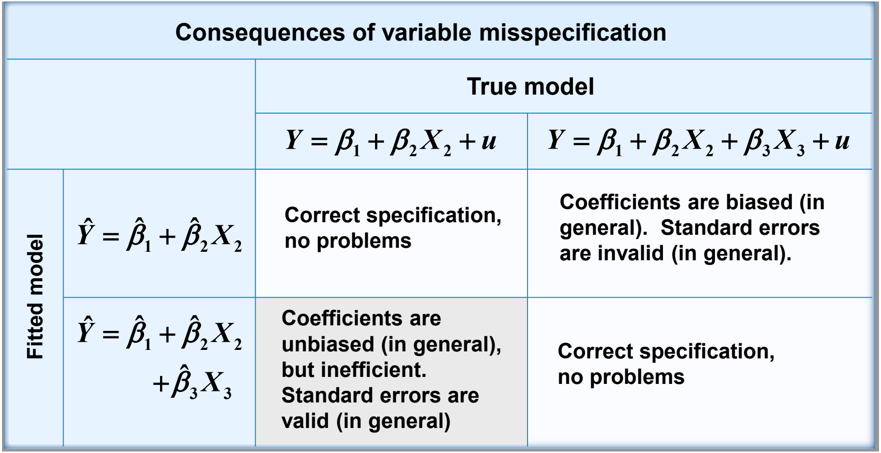
Variance:

However, the variance of $\hat{\beta}_2$ will be larger than it would have been if the correct simple regression had been run because it includes the factor $1/(1-\rho^2)$, where ρ is the correlation between X_2 and X_3.

$$\sigma^2_{\hat{\beta}_2} = \frac{\sigma_u^2}{\sum\left(X_{2i}-\bar{X}_2\right)} \times \frac{1}{1-\rho^2_{X_2,X_3}}$$

- The estimator b_2 using the multiple regression model will therefore be less efficient than the alternative using the simple regression model.
- The standard errors remain valid, because the model is formally correctly specified, but they will tend to be larger than those obtained in a simple regression, reflecting the loss of efficiency.
- Note that if X_2 and X_3 happen to be **uncorrelated**, there will be no loss of efficiency after all.

Omission of an Important Variable or Inclusion of an Irrelevant Variable



Consequences of variable misspecification

		True model	
		$Y = \beta_1 + \beta_2 X_2 + u$	$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$
Fitted model	$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2$	Correct specification, no problems	Coefficients are biased (in general). Standard errors are invalid (in general).
	$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$	Coefficients are unbiased (in general), but inefficient. Standard errors are valid (in general)	Correct specification, no problems

Figure: Details.

Section 6, Parameter stability

Parameter Stability Tests

- So far, we have estimated regressions such as

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

- We have implicitly assumed that the parameters (β_1, β_2 and β_3) are constant for the entire sample period.
- We can test this implicit assumption using parameter stability tests. The idea is essentially to split the data into sub-periods and then to estimate up to three models, for each of the sub-parts and for all the data and then to "compare" the *RSS* of the models.
- There are two types of test we can look at:
  - Chow test (analysis of variance test)
  - Predictive failure tests

The Chow Test I

- The steps involved are:
  1. Split the data into two sub-periods. Estimate the regression over the whole period and then for the two sub-periods separately (3 regressions). Obtain the RSS for each regression.
  2. The restricted regression is now the regression for the whole period while the "unrestricted regression" comes in two parts: for each of the sub-samples.
     We can thus form an F-test which is the difference between the *RSS*'s. The statistic is

$$test\ statistic = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2k}{k}$$

where:
RSS = RSS for whole sample
RSS_1 = RSS for sub-sample 1
RSS_2 = RSS for sub-sample 2

The Chow Test II

T = number of observations
$2k$ = number of regressors in the "unrestricted" regression (since it comes in two parts)
k = number of regressors in (each part of the) "unrestricted" regression

3. Perform the test. If the value of the test statistic is greater than the critical value from the F-distribution, which is an F(k, T-2k), then reject the null hypothesis that the parameters are stable over time.

A Chow Test Example

- Consider the following regression for the CAPM β (again) for the returns on Glaxo.
- Say that we are interested in estimating Beta for monthly data from 1981-1992. The model for each sub-period is
- $1981M1$–$1987M10$

$$\hat{r}_{gt} = 0.24 + 1.2 r_{Mt} \quad T = 82 \quad RSS_1 = 0.03555$$

- $1987M11$–$1992M12$

$$\hat{r}_{gt} = 0.68 + 1.53 r_{Mt} \quad T = 62 \quad RSS_2 = 0.00336$$

- $1981M1$–$1992M12$

$$\hat{r}_{gt} = 0.39 + 1.37 r_{Mt} \quad T = 144 \quad RSS = 0.0434$$

A Chow Test Example - Results

- The null hypothesis is

$$\mathrm{H}_0 : \ \alpha_1 = \alpha_2 \ \text{ and } \ \beta_1 = \beta_2$$

- The unrestricted model is the model where this restriction is not imposed

$$\begin{aligned} \textit{test statistic} &= \frac{0.0434 - (0.0355 + 0.00336)}{0.0355 + 0.00336} \times \frac{144 - 4}{2} \\ &= 7.698 \end{aligned}$$

- Compare with 5% F(2,140) = 3.06
- We reject H_0 at the 5% level and say that we reject the restriction that the coefficients are the same in the two periods.

The Predictive Failure Test I

- Problem with the Chow test is that we need to have enough data to do the regression on both sub-samples, i.e. $T_1 \gg k$, $T_2 \gg k$.

- An alternative formulation is the predictive failure test.

- What we do with the predictive failure test is estimate the regression over a "long" sub-period (i.e. most of the data) and then we predict values for the other period and compare the two.

To calculate the test:

- Run the regression for the whole period (the restricted regression) and obtain the RSS

The Predictive Failure Test II

- Run the regression for the "large" sub-period and obtain the corresponding RRS (called RSS_1).

$$test\ statistic = \frac{RSS - RSS_1}{RSS_1} \times \frac{T_1 - k}{T_2}$$

Once you have detected the "large" sub-period, the sample is divided in two:

- T_1 = number of observations in the first half of the sample
- T_2 = number of observations in the second half of the sample.

(That is, the RSS form the overall regression is computed on $T_1 + T_2$ observations.)

The test statistic will follow an $F(T_2, T_1 - k)$.

Predictive Failure Tests – An Example I

- We have the following models estimated:
  For the CAPM β on Glaxo.

- 1981M1–1992M12 (whole sample)

$$\hat{r}_{gt} = 0.39 + 1.37 r_{Mt} \qquad T = 144 \qquad RSS = 0.0434$$

- 1981M1–1990M12 ('long sub-sample')

$$\hat{r}_{gt} = 0.32 + 1.31 r_{Mt} \qquad T = 120 \qquad RSS_1 = 0.0420$$

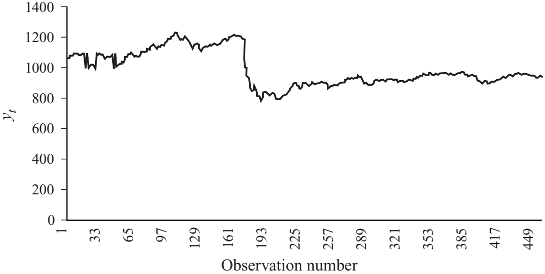
  How to apply the test?

$$\begin{aligned} \textit{test statistic} &= \frac{0.0434 - 0.0420}{0.0420} \times \frac{120-2}{24} \\ &= 0.164 \end{aligned}$$

- Compare the test statistic with an $F(24{,}118) = 1.66$ at the 5% level. So we do not reject the null hypothesis (about parameter's stability).

How do we decide the sub-parts to use?

- As a rule of thumb, we could use all or some of the following
  - Plot the dependent variable over time and split the data accordingly to any obvious structural changes in the series, e.g.



  - Split the data according to any known important historical events (e.g. stock market crash, new government elected)
  - Use all but the last few observations and do a predictive failure test on those.

Section 7, Error-in-variable models

Measurement Errors

- If there is measurement error in one or more of the explanatory variables, this will violate the assumption that the explanatory variables are non-stochastic
- Sometimes this is also known as the errors-in-variables problem
- Measurement errors can occur in a variety of circumstances, e.g.
  - Macroeconomic variables are almost always estimated quantities (GDP, inflation, and so on), as is most information contained in company accounts
  - Sometimes we cannot observe or obtain data on a variable we require and so we need to use a proxy variable – for instance, many models include expected quantities (e.g., expected inflation) but we cannot typically measure expectations.

Measurement Error in the Explanatory Variable(s) I

- Suppose that we wish to estimate a model containing just one explanatory variable, x_t:

$$y_t = \beta_1 + \beta_2 x_t + u_t$$

  where u_t is a disturbance term.

- Suppose further that x_t is measured with error so that instead of observing its true value, we observe a noisy version, $\tilde{x}$, that comprises the actual x_t plus some additional noise, v_t that is independent of x_t and u_t:

$$\tilde{x}_t = x_t + v_t$$

- Taking the first equation and substituting in for x_t from the second:

$$y_t = \beta_1 + \beta_2(\tilde{x}_t - v_t) + u_t$$

Measurement Error in the Explanatory Variable(s) II

- We can rewrite this equation by separately expressing the composite error term, $(u_t - \beta_2 v_t)$

$$y_t = \beta_1 + \beta_2 \tilde{x}_t + (u_t - \beta_2 v_t)$$

Measurement Error in the Explanatory Variable(s)

- It should be clear from this equation and the one for the explanatory variable measured with error, $\tilde{x}_t$ and the composite error term, $(u_t - \beta_2 v_t)$, are correlated since both depend on v_t
- Thus the requirement that the explanatory variables are non-stochastic does not hold
- This causes the parameters to be estimated inconsistently
- The size of the bias in the estimates will be a function of the variance of the noise in x_t as a proportion of the overall disturbance variance
- If β_2 is positive, the bias will be negative but if β_2 is negative, the bias will be positive
- So the parameter estimate will always be biased towards zero as a result of the measurement noise.

Measurement Error and Tests of the CAPM I

- The standard approach to testing the CAPM pioneered by Fama and MacBeth (1973) comprises two stages
- Since the betas are estimated at the first stage rather than being directly observable, they will surely contain measurement error
- The effect of this has sometimes been termed attenuation bias.
- Tests of the CAPM showed that the relationship between beta and returns was smaller than expected, and this is precisely what would happen as a result of measurement error
- Various approaches to solving this issue have been proposed, the most common of which is to use portfolio betas in place of individual betas
- An alternative approach (Shanken, 1992) is to modify the standard errors in the second stage regression to adjust directly for the measurement errors.

Measurement Error in the Explained Variable

- Measurement error in the explained variable is much less serious than in the explanatory variable(s)
- This is one of the motivations for the inclusion of the disturbance term in a regression model
- When the explained variable is measured with error, the disturbance term will in effect be a composite of the usual disturbance term and another source of noise from the measurement error
- Then the parameter estimates will still be consistent and unbiased and the usual formulae for calculating standard errors will still be appropriate
- The only consequence is that the additional noise means the standard errors will be enlarged relative to the situation where there was no measurement error in y.

Section 8, References

Slides

Disclaimer:

- Slides rearranged from Chris Brooks' slides from (Brooks, 2014) (copyrighted)

Bibliography I

[CO49] Donald Cochrane and Guy H Orcutt. ‘Application of least squares regression to relationships containing auto-correlated error terms’. In: *Journal of the American statistical association* 44.245 (1949), pp. 32–61.

[GQ65] Stephen M Goldfeld and Richard E Quandt. ‘Some tests for homoscedasticity’. In: *Journal of the American statistical Association* 60.310 (1965), pp. 539–547.

[Ram69] James Bernard Ramsey. ‘Tests for specification errors in classical linear least-squares regression analysis’. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 31.2 (1969), pp. 350–371.

Bibliography II

[HM78] David F Hendry and Grayham E Mizon. ‘Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England’. In: *The Economic Journal* 88.351 (1978), pp. 549–563.

[JB80] Carlos M Jarque and Anil K Bera. ‘Efficient tests for normality, homoscedasticity and serial independence of regression residuals’. In: *Economics letters* 6.3 (1980), pp. 255–259.

[Whi80] Halbert White. ‘A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity’. In: *Econometrica: journal of the Econometric Society* (1980), pp. 817–838.

[DW92] James Durbin and Geoffrey S Watson. ‘Testing for serial correlation in least squares regression. I’. In: *Breakthroughs in Statistics*. Springer, 1992, pp. 237–259.

Bibliography III

[Sha92] Jay Shanken. ‘On the estimation of beta-pricing models’. In: *The review of financial studies* 5.1 (1992), pp. 1–33.

[Bro14] Chris Brooks. *Introductory Econometrics for Finance*. 3rd ed. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139540872.