Slide 8.1

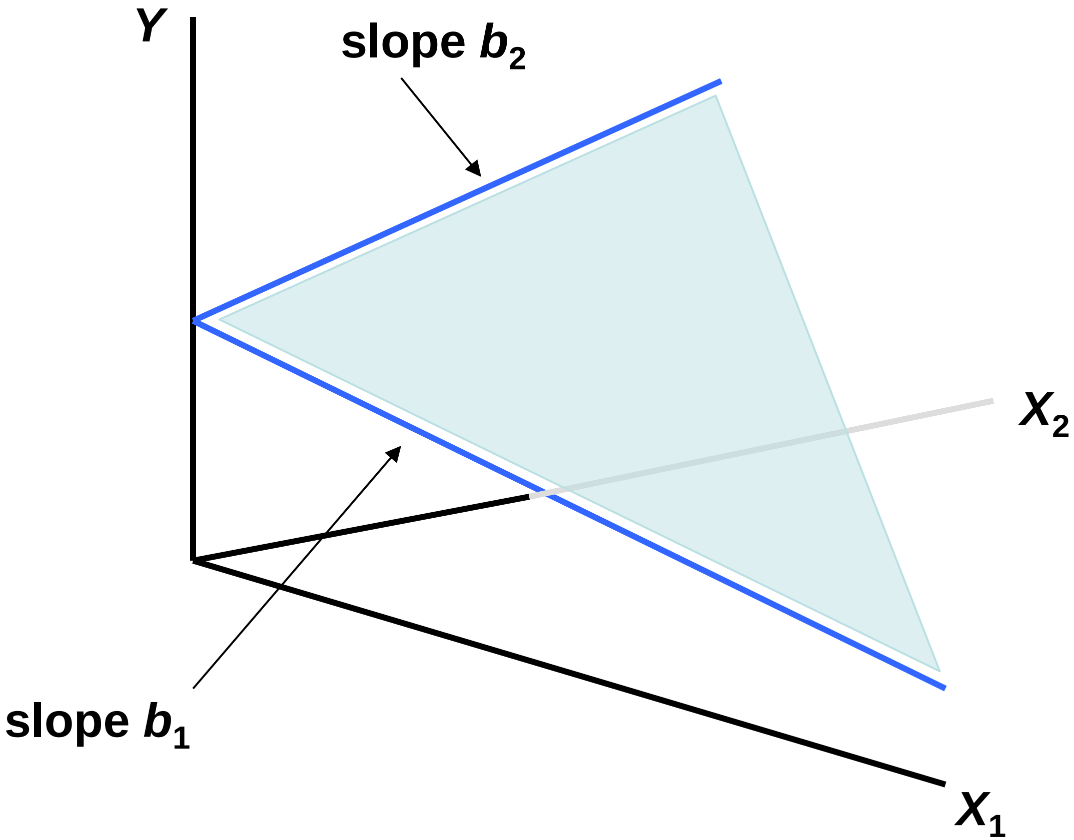
Chapter 8: Multiple regression

- We can extend the regression model to allow several explanatory variables. The sample regression equation becomes

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k + e$$

Picture of the regression model

- With two *X* variables: $Y = b_0 + b_1 X_1 + b_2 X_2 + e$



Obtaining the regression equation

- The principles are the same: minimise the sum of squared errors (vertical distances from the regression plane)
- The calculations are more complex - use a computer

Slide 8.4

Example: import demand equation

Year	Imports	GDP	GDP deflator	Price of imports	RPI all items
1973	18.8	74.0	24.6	21.5	25.1
1974	27.0	83.8	28.7	31.3	29.1
1975	28.7	105.9	35.7	35.6	36.1
:	:	:	:	:	:
2003	314.8	1110.3	195.6	106.7	191.7
2004	333.7	1176.5	201.0	106.2	197.4
2005	366.5	1224.7	205.4	110.7	202.9

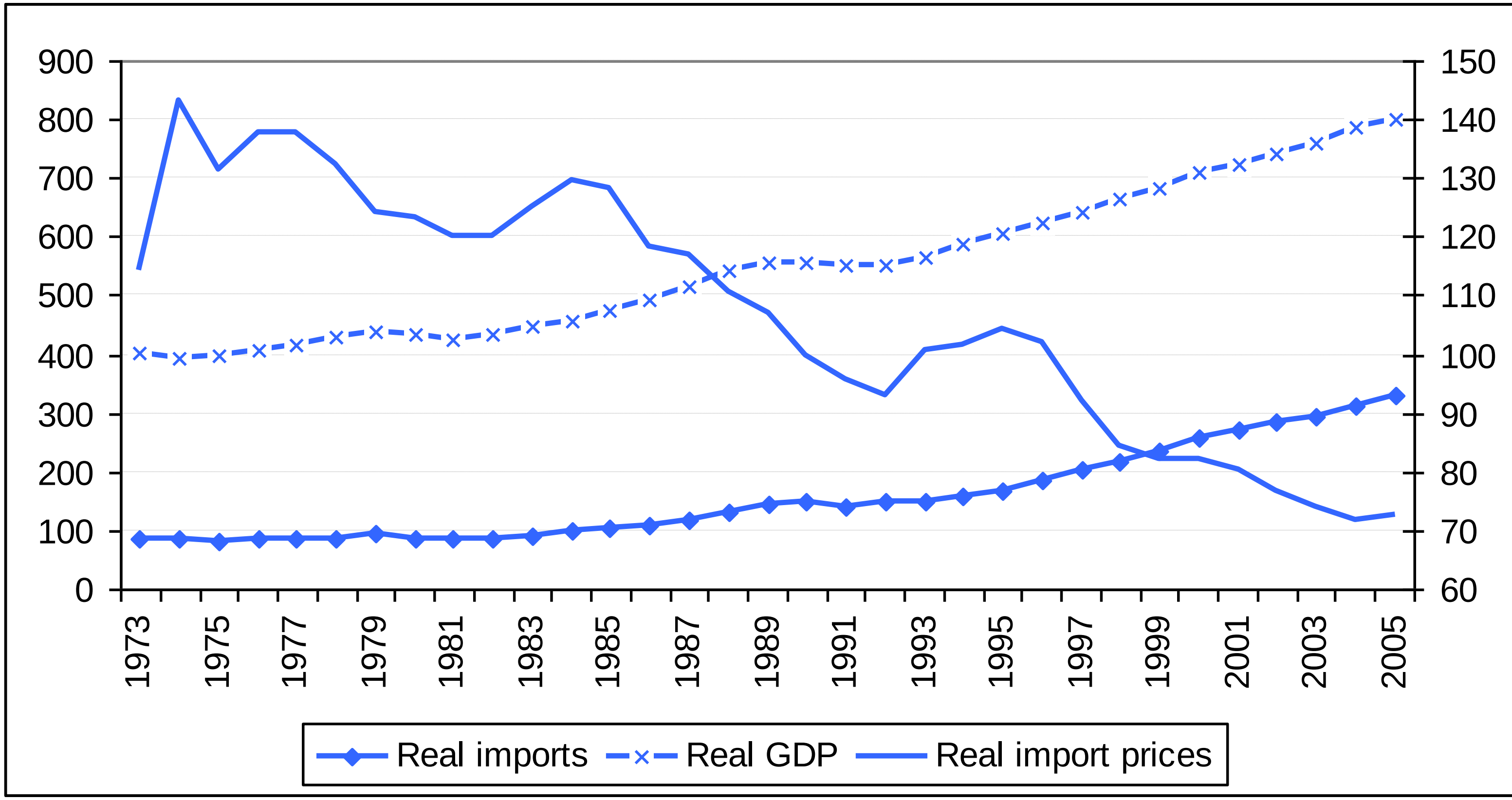
Slide 8.5

Data transformed to real values

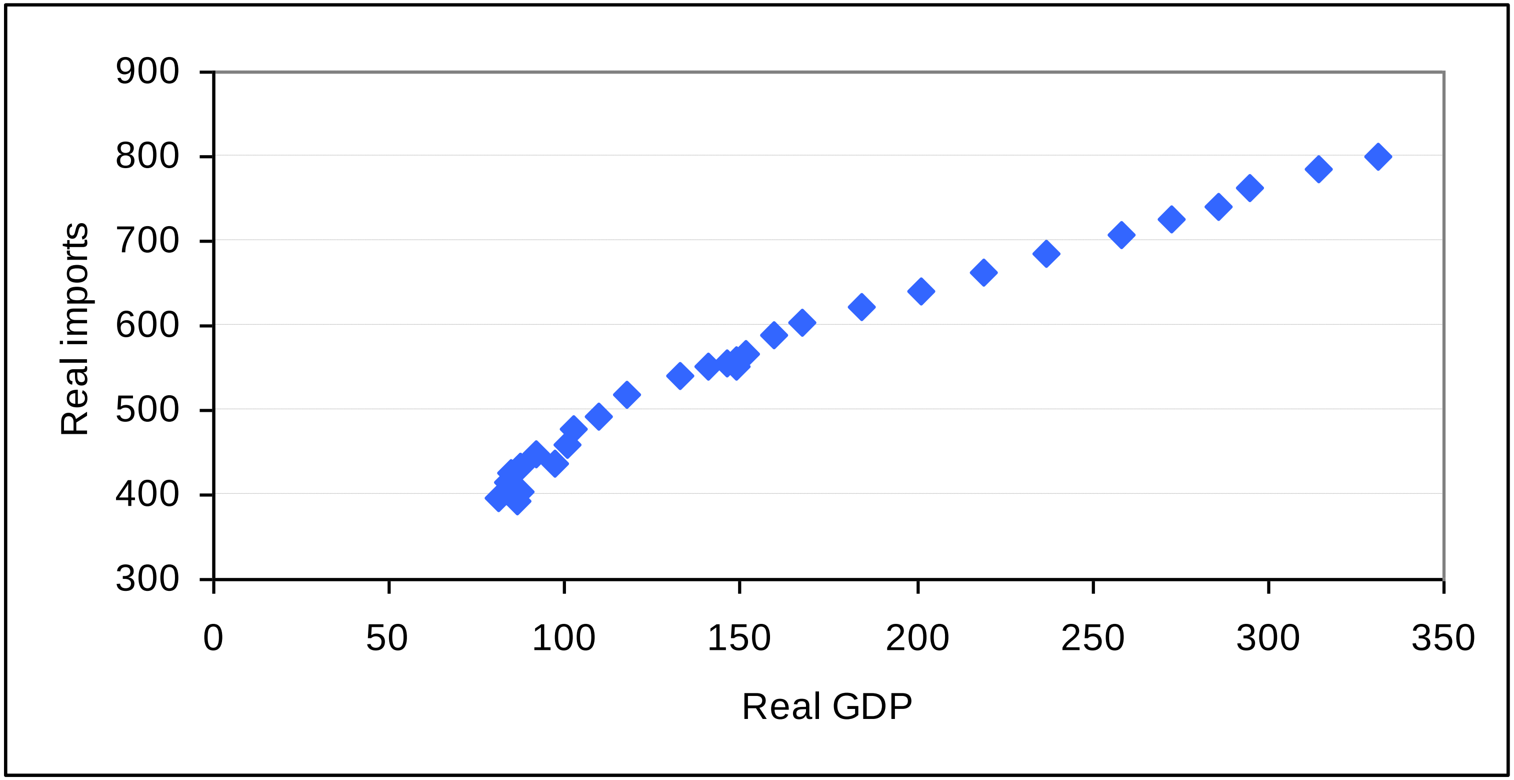
Year	Real imports	Real GDP	Real import prices
1973	87.4	403.4	114.2
1974	86.3	391.6	143.4
1975	80.6	397.8	131.5
:	:	:	:
2003	295	761.2	74.2
2004	314.2	784.9	71.7
2005	331.1	799.6	72.7

Slide 8.6

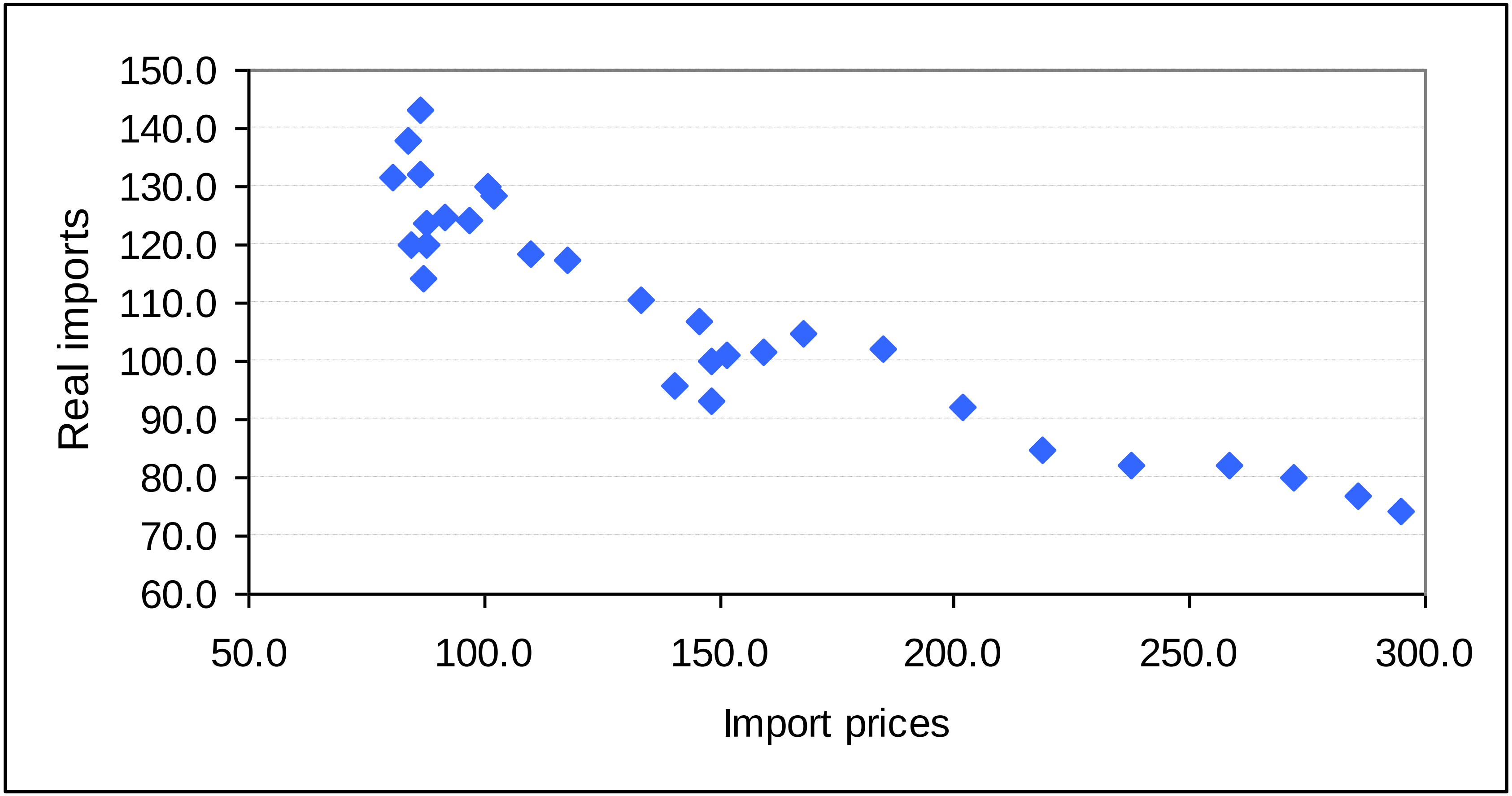
Time series chart of data



XY chart: imports and GDP



XY chart: imports and prices



Regression results (via Excel)

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.98					
R Square	0.96					
Adjusted R Square	0.96					
Standard Error	13.24					
Observations	31					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	129031.05	64515.52	368.23	7.82025E-21	
Residual	28	4905.70	175.20			
Total	30	133936.75				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-172.61	73.33	-2.35	0.03	-322.83	-22.39
Real GDP	0.59	0.06	9.12	0.00	0.45	0.72
Real import prices	0.05	0.37	0.13	0.90	-0.70	0.79

Interpreting the coefficients

- Effect of GDP on imports: 0.59
- Better to calculate the elasticity:

$$\eta_{gdp} = b_1 \times \frac{\overline{gdp}}{\overline{m}} = 0.59 \times \frac{536.4}{146.3} = 2.16$$

- A 1% rise in GDP leads to a 2% (approx) increase in imports
- The price elasticity is 0.04, by a similar calculation

Significance tests of the coefficients

- For GDP, t = 9.12, highly significant (t^*_{28} = 2.048 or 1.701 for a one tail test)
- For price, t = 0.13, not significant
- The price effect is the wrong sign, small and statistically not significant

Goodness of fit

- $R^2 = 0.96$. 96% of the variation in imports is explained by variation in GDP and prices

- Testing H_0: $R^2 = 0$ we obtain

$$F = \frac{RSS/k}{ESS/(n-k-1)} = \frac{129{,}031.05/2}{4905.70/(31-2-1)} = 368.23$$

which is highly significant ($F^*_{2,28} = 3.34$)

An equivalent hypothesis

- Testing H_0: $R^2 = 0$ is equivalent to testing that all the slope coefficients are zero, i.e.

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_0: \beta_1 \neq \beta_2 \neq 0$$

- The null implies *neither* GDP *nor* price influences imports. As we have seen, this is rejected.

Prediction

- Predicting imports for 2002–3 we obtain:
  - 2004: $\hat{m} = -172.61 + 0.59 \times 784.9 + 0.05 \times 71.7 = 290.0$
  - 2005: $\hat{m} = -172.61 + 0.59 \times 799.6 + 0.05 \times 72.7 = 298.6$
- The error from the actual values is around 12%

Year	Actual	Forecast	Error
2004	314.2	290.0	24.2
2005	331.1	298.6	32.5

Estimating in logs

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.99					
R Square	0.98					
Adjusted R Square	0.98					
Standard Error	0.05					
Observations	31					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	5.31	2.65	901.43	3.82835E-26	
Residual	28	0.08246	0.00			
Total	30	5.39				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-3.60	1.65	-2.17	0.04	-6.98	-0.21
ln GDP	1.66	0.15	11.31	0.00	1.36	1.97
ln import prices	-0.41	0.16	-2.56	0.02	-0.74	-0.08

Interpreting the result

- GDP and price elasticities are 1.66 and -0.48 respectively
- Both are statistically significant
- Predicting for 2004 gives

$$\ln \hat{m} = -3.60 + 1.66 \times 6.67 - 0.41 \times 4.27 = 5.73$$

- taking the anti-log gives $e^{5.73} = 308.2$

Predictions

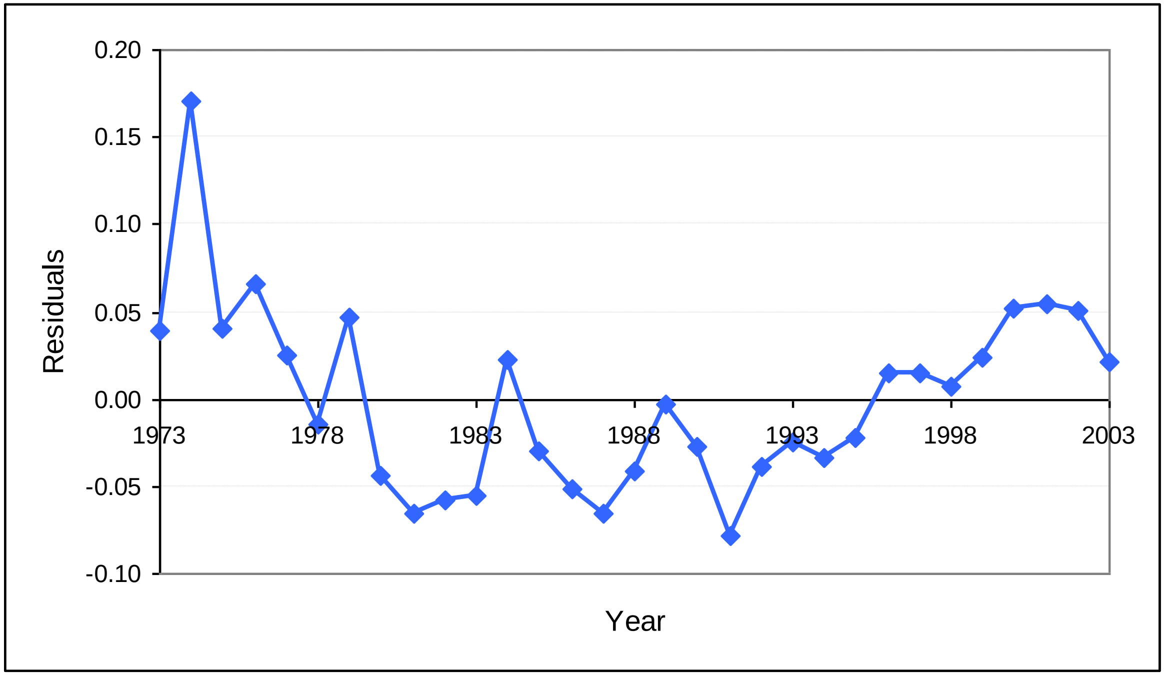
- The prediction errors are now smaller: 1.9% and 4.8% in the two years

Year	Actual	Fitted	Error	% error
2004	314.2	308.2	6.0	1.9
2005	331.1	316.0	15.1	4.8

Slide 8.18

Autocorrelation

- The pattern of errors (over time) should be random



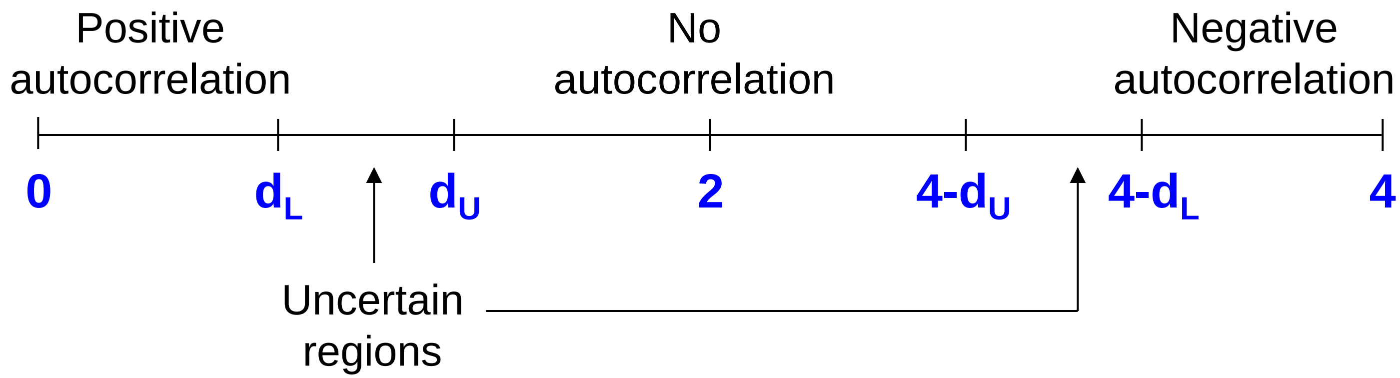
Errors from log model

Slide 8.19

The Durbin – Watson statistic

- Provides a test for autocorrelation

$$\mathrm{DW} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$



The Durbin – Watson statistic (continued)

$$DW = \frac{0.0705}{0.0825} = 0.855$$

	e_t	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
1973	0.0396	0.0000	0.0396		0.0016
1974	0.1703	0.0396	0.1308	0.0171	0.0290
1975	0.0401	0.1703	-0.1302	0.0170	0.0016
:	:	:	:	:	:
2002	0.0509	0.0548	-0.0039	0.0000	0.0026
2003	0.0215	0.0509	-0.0294	0.0009	0.0005
Totals				0.0705	0.0825

- For $n = 30$, $k = 2$, $d_L = 1.284$, $d_U = 1.567$, hence positive autocorrelation present

Consequences of autocorrelation

- Forecasts not optimal (too low in this case)
- Possible spurious regression (especially when variables are trended)
- *t* and *F* statistics biased upwards
- A warning to investigate further

Restricted and unrestricted models

- Restricted model (real price):
  - $\ln m = b_0 + b_1 \ln gdp + b_2 \ln p_m + e$
- Unrestricted model(nominal prices):
  - $\ln m = c_0 + c_1 \ln gdp + c_2 \ln P_M + c_3 \ln P + e$
- Test H_0: $c_2 = -c_3$

Slide 8.23

Restricted and unrestricted models (continued)

- Unrestricted model *must* fit better
- But if H_0 is true, restricted model should fit almost as well. Hence compare ESS_R with ESS_U
- Test statistic is:

$$F = \frac{(\mathrm{ESS_R} - \mathrm{ESS_U})/q}{\mathrm{ESS_U}/(n-k-1)}$$

Slide 8.24

Restricted and unrestricted models (continued)

- The unrestricted model is estimated as:

$$\ln m_t = -8.77 + 2.31 \ln gdp_t - 0.20 \ln P_{Mt-1} + 0.02 \ln P_{t-1} + e_t$$

with $ESS_U = 0.0272$. Hence we obtain:

$$F = \frac{(ESS_R - ESS_U)/1}{ESS_U/(31-3-1)} = \frac{(0.08246 - 0.02720)/1}{0.02720/(31-3-1)} = 54.85$$

- $> F^*_{1,28} = 4.21$, so H_0 is rejected, perhaps surprisingly.

Summary

- Multiple regression extends the two variable model.
- Similar principles, different calculations
- Data transformations, e.g. logs, can be useful
- The adequacy of the model can be assessed by its forecasts and by checking for autocorrelation (amongst other things)
- Unrestricted and restricted models can be compared using an *F* test