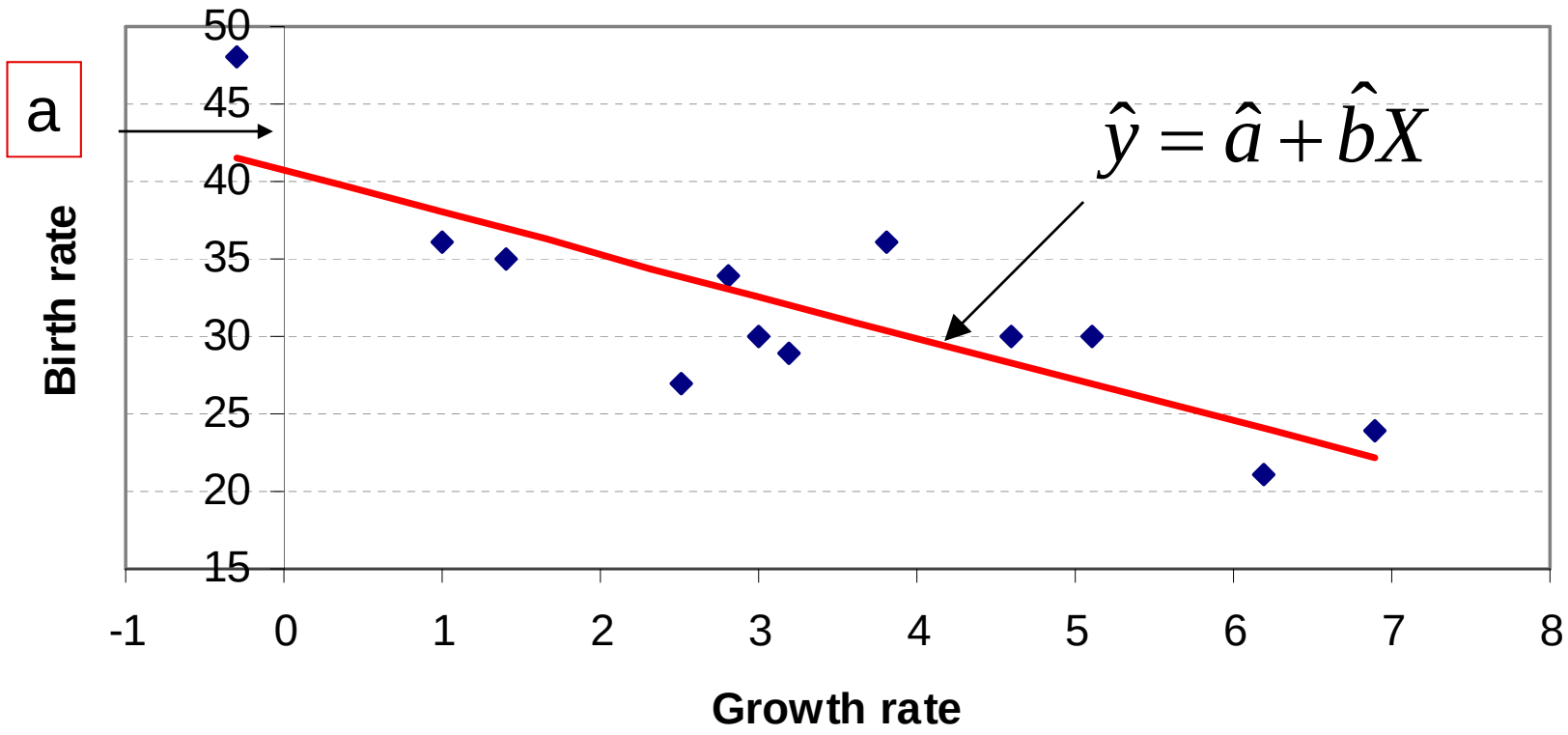
Regression (2)

Lecture 7

Regression (2)

- Recall that regression analysis allows us to
  - measure the effect of *X* upon *Y* and
  - allows several explanatory variables to influence *Y*

Regression Line: the Line of "Best Fit"



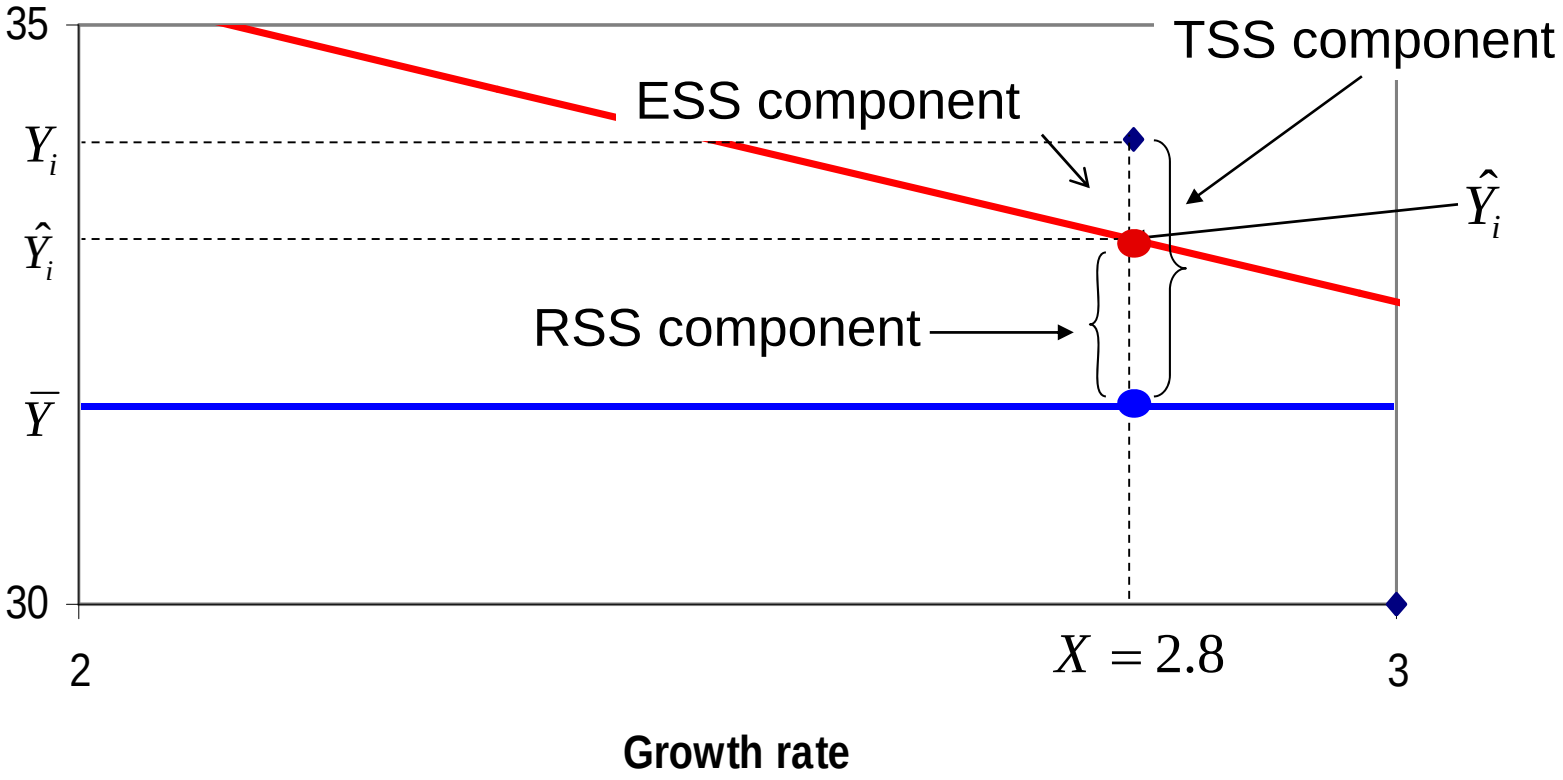
Measuring the Goodness of Fit

- Use the coefficient of determination, R^2

$$R^2 = \frac{RSS}{TSS}$$

- $0 \leq R^2 \leq 1$
- RSS: regression sum of squares
  TSS: total sum of squares
- NB: notation varies in different textbooks

The Component Parts of R^2



Calculating Sums of Squares

- TSS = RSS + ESS

$$TSS = \sum\left(Y - \bar{Y}\right)^2 = \sum Y^2 - n\bar{Y}^2 = 12,564 - 12 \times 31.67^2 = 530.67$$

$$ESS = \sum\left(Y - \hat{Y}\right)^2 = \sum Y^2 - a\sum Y - b\sum XY$$

$$= 12,564 - 40.71 \times 380 - \left(-2.7\right) \times 1,139.7 = 170.75$$

$$RSS = 530.67 - 170.75 = 359.92$$

- Hence

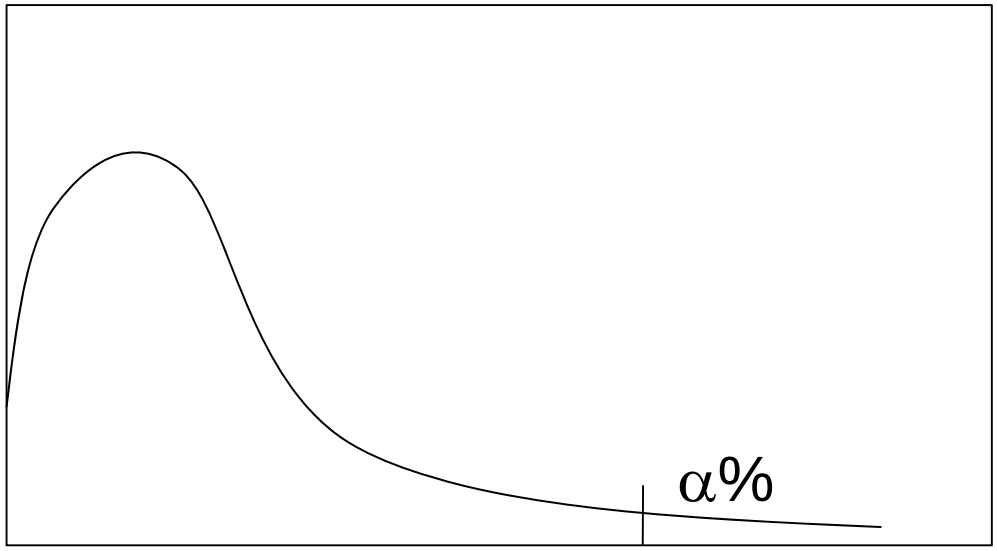
$$R^2 = \frac{359.92}{530.67} = 0.678$$

  - Our model explains 67.8% of the variation in birth rates
  - Note that $r^2=R^2$ in the case of 1 indep. var

Testing the Goodness of Fit

- Is the model any good?
  - Is it better than birth-rate=a+e?
- H_0: b=0

$$F = \frac{R^2 / k}{(1 - R^2)/(n - k - 1)} \sim F_{k,n-k-1,\alpha}$$



F distribution is
- asymmetric
- positive

Critical Values of the F-Distribution (5%)

	1	2	3	4
..				
2	18.513	19.00	19.164	19.247
3	10.128	9.5521	9.2766	9.1172
..				
10	4.9646	4.1028	3.7083	3.4780
....				

F-test

- In the birth & growth rate example, R^2=0.678, n=12, k=1 so the F statistic=21.08
- Critical value at 5% level with 1 and 10 degrees of freedom is 4.9646
- Therefore we reject the null and conclude that the model *does* have explanatory power
- Alternative formula for F test based on error sum of squares

Multiple Regression

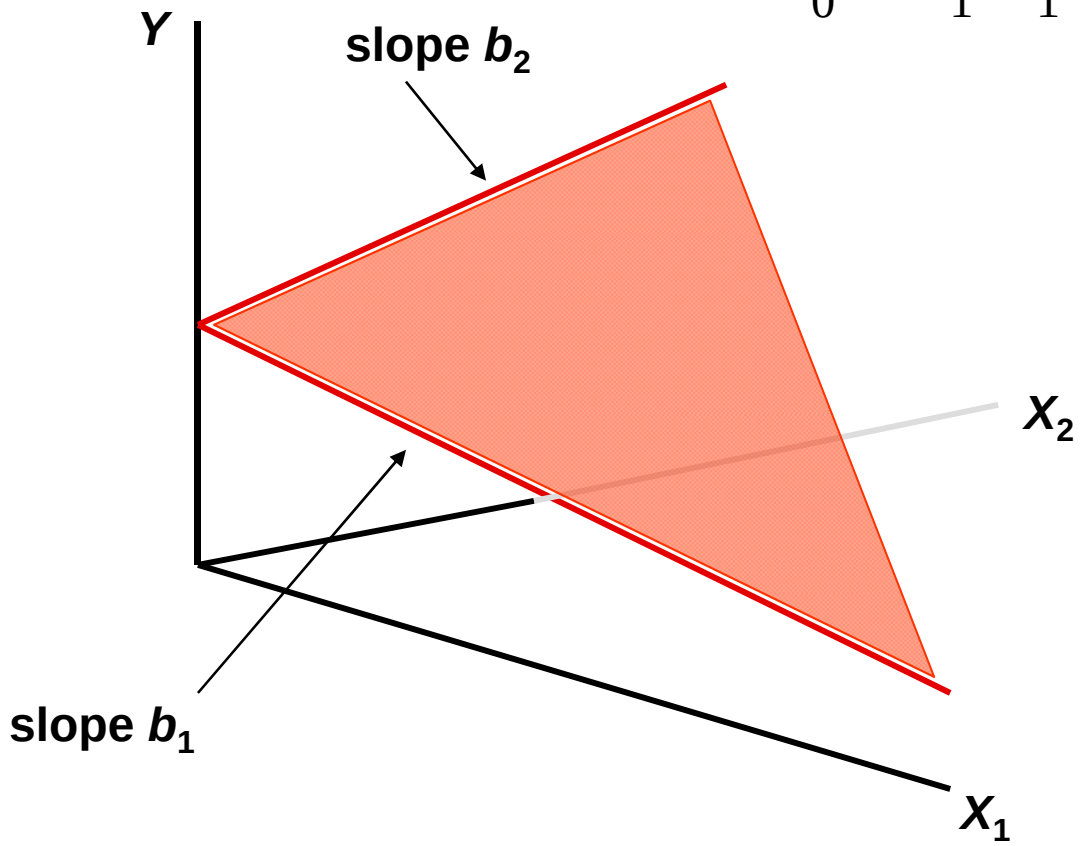
- We can extend the regression model to allow several explanatory variables. The sample regression equation becomes

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

Picture of the Regression Model

- With two *explanatory* 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 – need to use a computer

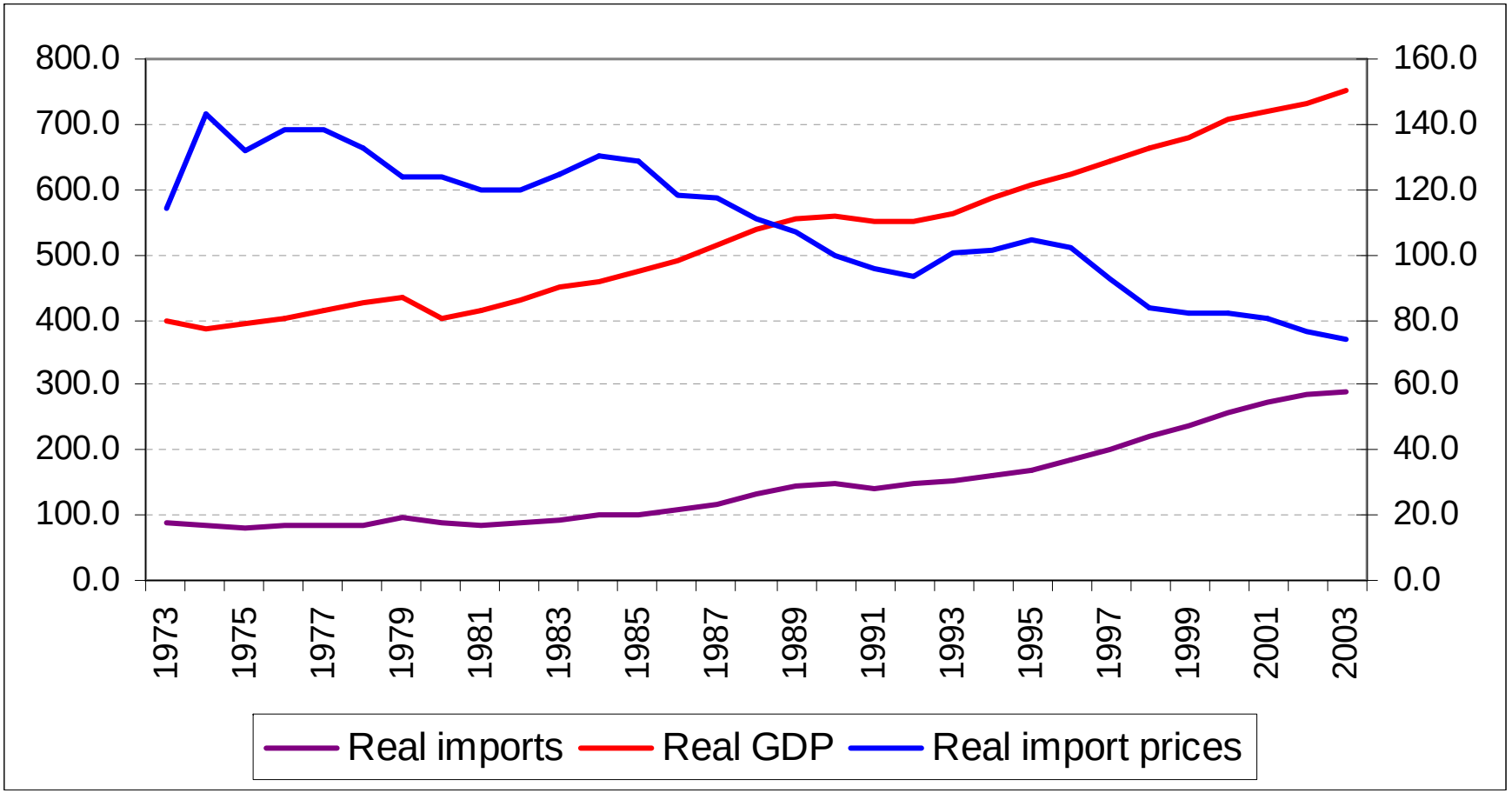
Example: Import Demand Equation

Year	Imports	GDP	GDP deflator	Price of imports	RPI all items
1973	18.8	74.0	24.8	21.5	25.1
1974	27.0	83.7	28.9	31.3	29.1
1975	28.7	105.8	35.8	35.6	36.1
:	:	:	:	:	:
2001	299.3	994.0	184.6	110.2	183.2
2002	304.7	1043.3	190.5	107.2	186.3
2003	308.4	1099.4	196.1	106.6	191.7

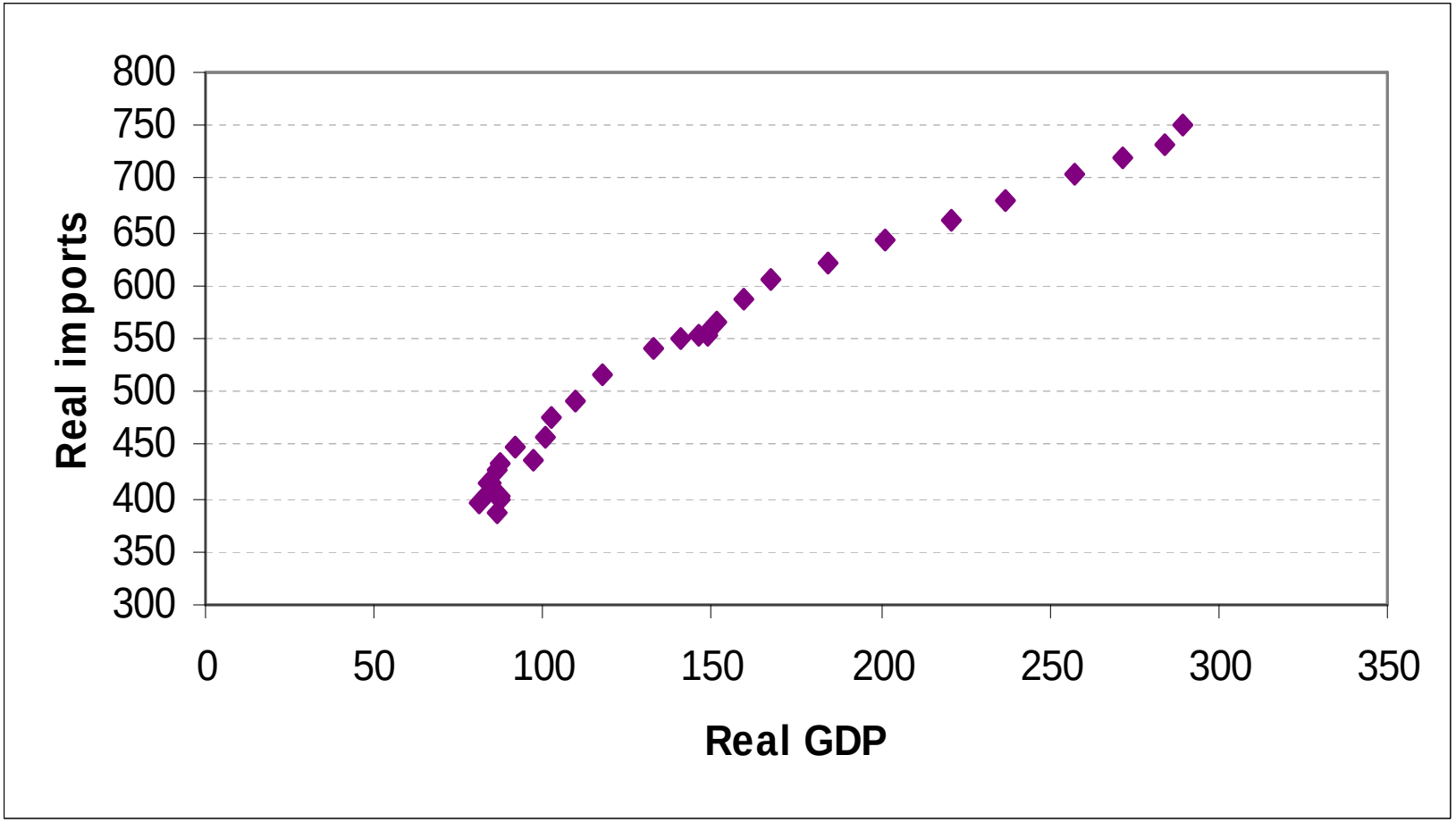
Import Demand Data Transformed to Real Values

Year	Real imports	Real GDP	Real import prices
1973	87.4	399.5	114.2
1974	86.3	387.8	143.4
1975	80.6	395.7	131.5
:	:	:	:
2001	271.6	721.0	80.2
2002	284.2	733.3	76.7
2003	289.3	750.7	74.1

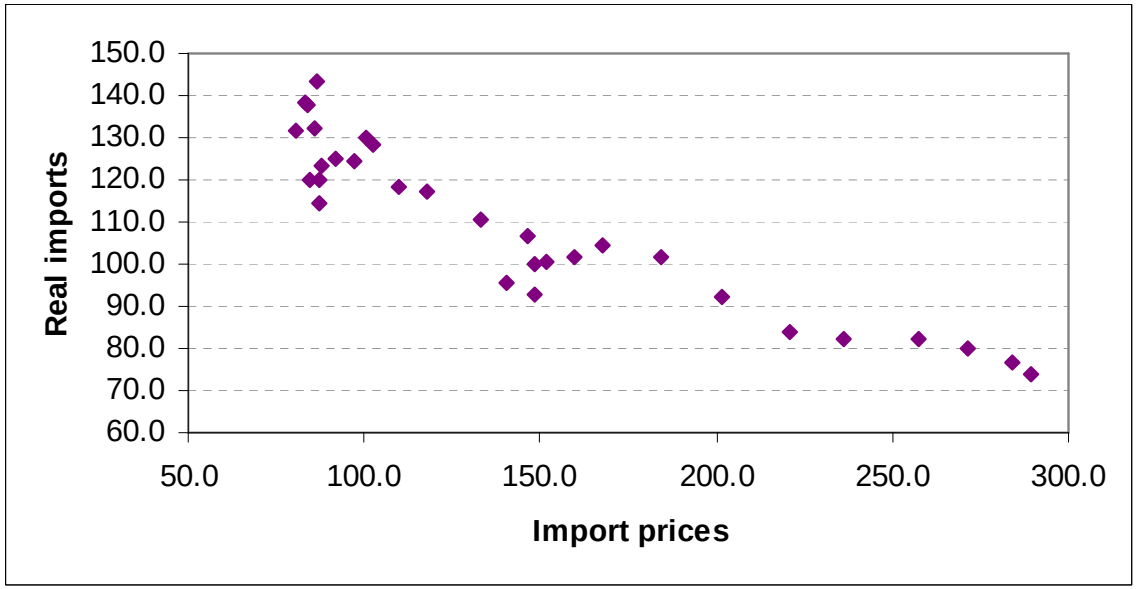
Time Series Chart of Data



XY Chart: Imports and GDP



XY Chart: Imports and Prices



Regression Results (using Excel)

Imports=$a+b_1$RGDP +b_2RPM

Regression Statistics	
Multiple R	0.98
R Square	0.95
Adjusted R Square	0.95
Standard Error	12.69
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	85209.41	42604.70	264.64	5.2E-18
Residual	26	4185.82	160.99		
Total	28	89395.23			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-103.22	67.96	-1.52	0.14	-242.92	36.48
Real GDP	0.50	0.06	8.30	0.00	0.38	0.63
Real import prices	-0.19	0.34	-0.57	0.58	-0.89	0.50

Interpreting the Coefficients

- Effect of GDP on imports: 0.5
- Better to calculate the elasticity:

$$\eta_{gdp} = b_1 \times \frac{\overline{gdp}}{\overline{m}} = 0.5 \times \frac{518.9}{136.4} = 1.90$$

- A 1% rise in GDP leads to a 2% (approx.) increase in imports
- The price elasticity is -0.16, by a similar calculation

Significance Tests of the Coefficients

- For GDP, $t = 8.3$, highly significant ($t^*_{26} = 2.056$)
- For price, $t = -0.57$, not significant
- The price effect is both small and statistically not significant

Goodness of Fit

- R^2 = 0.95. 95% of the variation in imports is explained by variation in GDP and prices
- Testing H_0: R^2 = 0 we obtain

$$F = \frac{R^2/k}{(1-R^2/(n-k-1)} = \frac{0.95/2}{0.05/(29-2-1)} = 264.6$$

which is highly significant ($F^*_{2,26,\,0.05}$ = 3.37)

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

imports=-103.2+0.5RGDP-0.19RPM

- Predicting imports for 2002 and 2003:
  - 2002: -103.2 + 0.5 $\times$ 733.3 - 0.19 $\times$ 76.7 = 250.9
  - 2003: -103.2 + 0.5 $\times$ 750.7 - 0.19 $\times$ 74.1 = 260.2
- The error from the actual values is around 12%

Year	Actual	Forecast	Error
2002	284.2	250.9	33.3
2003	289.3	260.2	29.1

Estimating in Logs

Regression Statistics	
Multiple R	0.99
R Square	0.98
Adjusted R Square	0.98
Standard Error	0.06
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	4.02	2.01	638.10	8.01E-23
Residual	26	0.08	0.00		
Total	28	4.10			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2.52	1.61	-1.57	0.13	-5.83	0.79
ln GDP	1.54	0.14	10.92	0.00	1.25	1.83
ln import prices	-0.48	0.16	-2.95	0.01	-0.81	-0.15

Interpreting the Results

- Taking logs on both sides means we can interpret coefficients as elasticities
  - GDP and price elasticities are 1.54 and -0.48 respectively
- Both are statistically significant (t-ratios>|2|)
- Predicting log(imports) for 2002 gives

  $-2.52 + 1.54 \times 6.60 - 0.48 \times 4.34 = 5.58$

  taking the anti-log gives $e^{5.58} = 264.1$

Prediction

- The prediction errors are now smaller: 8% and 4% in the two years

Year	Actual	Fitted	Error	% error
2002	284.2	264.1	20.1	7.6
2003	289.3	278.4	10.9	3.9

- Non-linear transformations often improve the model

Summary

- Multiple regression extends the two variable model.
- Similar principles, different calculations
- Data transformations, e.g. logs, can be useful