Chapter 4

Further development and analysis of the classical linear regression model

Generalising the Simple Model to Multiple Linear Regression

- Before, we have used the model

$$y_t = \alpha + \beta x_t + u_t \quad t = 1,2,...,T$$

- But what if our dependent (y) variable depends on more than one independent variable?
  For example the number of cars sold might plausibly depend on
  1. the price of cars
  2. the price of public transport
  3. the price of petrol
  4. the extent of the public's concern about global warming
- Similarly, stock returns might depend on several factors.
- Having just one independent variable is no good in this case - we want to have more than one x variable. It is very easy to generalise the simple model to one with $k - 1$ regressors (independent variables).

Multiple Regression and the Constant Term

- Now we write

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + ... + \beta_k x_{kt} + u_t, \; t=1,2,..., \; T$$

- Where is x_1? It is the constant term. In fact the constant term is usually represented by a column of ones of length T:

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

β_1 is the coefficient attached to the constant term (which we called α before).

Different Ways of Expressing the Multiple Linear Regression Model

- We could write out a separate equation for every value of t:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \cdots + \beta_k x_{k1} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \cdots + \beta_k x_{k2} + u_2 \\ \cdots & \quad \cdots \quad \cdots \quad \cdots \\ y_T &= \beta_1 + \beta_2 x_{2T} + \beta_3 x_{3T} + \cdots + \beta_k x_{kT} + u_T \end{aligned}$$

- We can write this in matrix form

$$y = X\beta + u$$

where: y is $T \times 1$
X is $T \times k$
β is $k \times 1$
u is $T \times 1$

Inside the Matrices of the Multiple Linear Regression Model

- e.g. if k is 2, we have 2 regressors, one of which is a column of ones:

$$\underset{T \times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}} = \underset{T \times 2}{\begin{bmatrix} 1 & x_{21} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{2T} \end{bmatrix}} \underset{2 \times 1}{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}} + \underset{T \times 1}{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}}$$

- Notice that the matrices written in this way are conformable.

How Do We Calculate the Parameters (the β) in this Generalised Case?

- Previously, we took the residual sum of squares, and minimised it w.r.t. α and β.
- In the matrix notation, we have

$$\hat{u} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_T \end{bmatrix}$$

- The RSS would be given by

$$\hat{u}'\hat{u} = [\hat{u}_1 \ \hat{u}_2 \ \cdots \hat{u}_T] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_T \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_T^2 = \sum \hat{u}_t^2$$

The OLS Estimator for the Multiple Regression Model

- In order to obtain the parameter estimates, β_1, β_2,..., β_k, we would minimise the RSS with respect to all the βs.
- It can be shown that

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = (X'X)^{-1}X'y$$

Calculating the Standard Errors for the Multiple Regression Model

- Check the dimensions: $\hat{\beta}$ is $k \times 1$ as required.
- But how do we calculate the standard errors of the coefficient estimates?
- Previously, to estimate the variance of the errors, σ^2, we used $s^2 = \frac{\sum \hat{u}^2}{T-2}$.
- Now using the matrix notation, we use

$$s^2 = \frac{\hat{u}'\hat{u}}{T-k}$$

- where k = number of regressors. It can be proved that the OLS estimator of the variance of $\hat{\beta}$ is given by the diagonal elements of $s^2(X'X)^{-1}$, so that the variance of $\hat{\beta}_1$ is the first element, the variance of is the second element, and ... , and the variance of $\hat{\beta}_k$ is the k^{th} diagonal element.

Calculating Parameter and Standard Error Estimates for Multiple Regression Models: An Example

- Example: The following model with k=3 is estimated over 15 observations:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

and the following data have been calculated from the original X's.

$$(X'X)^{-1} = \begin{bmatrix} 2.0 & 3.5 & -1.0 \\ 3.5 & 1.0 & 6.5 \\ -1.0 & 6.5 & 4.3 \end{bmatrix}, \quad (X'y) = \begin{bmatrix} -3.0 \\ 2.2 \\ 0.6 \end{bmatrix}, \quad \hat{u}'\hat{u} = 10.$$

Calculating Parameter and Standard Error Estimates for Multiple Regression Models: An Example (Cont'd)

Calculate the coefficient estimates and their standard errors.

- To calculate the coefficients, just multiply the matrix by the vector to obtain $(X'X)^{-1}X'y$.
- To calculate the standard errors, we need an estimate of σ^2.

$$s^2 = \frac{RSS}{T-k} = \frac{10.96}{15-3} = 0.91$$

- The variance-covariance matrix of $\hat{\beta}$ is given by

$$s^2(X'X)^{-1} = 0.91(X'X)^{-1} = \begin{bmatrix} 1.82 & 3.19 & -0.91 \\ 3.19 & 0.91 & 5.92 \\ -0.91 & 5.92 & 3.91 \end{bmatrix}$$

Calculating Parameter and Standard Error Estimates for Multiple Regression Models: An Example (Cont'd)

- The variances are on the leading diagonal:

$$\begin{array}{lcl} \text{var}(\hat{\beta}_1) = 1.82 & & SE(\hat{\beta}_1) = 1.35 \\ \text{var}(\hat{\beta}_2) = 0.91 & \Leftrightarrow & SE(\hat{\beta}_2) = 0.95 \\ \text{var}(\hat{\beta}_3) = 3.91 & & SE(\hat{\beta}_3) = 1.98 \end{array}$$

- We write:

$$\begin{array}{lccc} \hat{y} & = 1.10 - 4.40x_2 & + 19.88x_3 \\ & (1.35)\ (0.96) & (1.98) \end{array}$$

Testing Multiple Hypotheses: The F-test

- We used the t-test to test single hypotheses, i.e. hypotheses involving only one coefficient. But what if we want to test more than one coefficient simultaneously?
- We do this using the F-test. The F-test involves estimating 2 regressions.
- The unrestricted regression is the one in which the coefficients are freely determined by the data, as we have done before.
- The restricted regression is the one in which the coefficients are restricted, i.e. the restrictions are imposed on some βs.

The F-test: Restricted and Unrestricted Regressions

- Example

  The general regression is

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$

- We want to test the restriction that $\beta_3 + \beta_4 = 1$ (we have some hypothesis from theory which suggests that this would be an interesting hypothesis to study). The unrestricted regression is (13) above, but what is the restricted regression?

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad \text{s.t. } \beta_3 + \beta_4 = 1$$

- We substitute the restriction ($\beta_3 + \beta_4 = 1$) into the regression so that it is automatically imposed on the data.

$$\beta_3 + \beta_4 = 1 \Rightarrow \beta_4 = 1 - \beta_3$$

The F-test: Forming the Restricted Regression

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + (1 - \beta_3) x_{4t} + u_t \\
y_t &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + x_{4t} - \beta_3 x_{4t} + u_t
\end{aligned}
$$

- Gather terms in β's together and rearrange

$$
(y_t - x_{4t}) = \beta_1 + \beta_2 x_{2t} + \beta_3 (x_{3t} - x_{4t}) + u_t
$$

- This is the restricted regression. We actually estimate it by creating two new variables, call them, say, P_t and Q_t.

$$
\begin{aligned}
P_t &= y_t - x_{4t} \\
Q_t &= x_{3t} - x_{4t}
\end{aligned}
$$

So $P_t = \beta_1 + \beta_2 x_{2t} + \beta_3 Q_t + u_t$ is the restricted regression we actually estimate.

Calculating the F-Test Statistic

- The test statistic is given by

$$test\ statistic = \frac{RRSS - URSS}{URSS} \times \frac{T - k}{m}$$

where $URSS$ = *RSS from unrestricted regression*

$RRSS$ = *RSS from restricted regression*

m = number of restrictions

T = number of observations

k = number of regressors in unrestricted regression including a constant in the unrestricted regression (or the total number of parameters to be estimated).

The F-Distribution

- The test statistic follows the F-distribution, which has 2 d.f. parameters.
- The value of the degrees of freedom parameters are m and (T-k) respectively (the order of the d.f. parameters is important).
- The appropriate critical value will be in column m, row (T-k).
- The F-distribution has only positive values and is not symmetrical. We therefore only reject the null if the test statistic $>$ critical F-value.

Determining the Number of Restrictions in an F-test

- Examples :

H_0 : *hypothesis*	*No. of restrictions*, m
$\beta_1 + \beta_2 = 2$	1
$\beta_2 = 1$ and $\beta_3 = -1$	2
$\beta_2 = 0, \beta_3 = 0$ and $\beta_4 = 0$	3

- If the model is $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_{4t} x_{4t} + u_t$, then the null hypothesis
  $H_0 : \beta_2 = 0$, and $\beta_3 = 0$ and $\beta_4 = 0$ is tested by the regression F-statistic. It tests the null hypothesis that all of the coefficients except the intercept coefficient are zero.
- Note the form of the alternative hypothesis for all tests when more than one restriction is involved: $H_1 : \beta_2 \neq 0$, or $\beta_3 \neq 0$ or $\beta_4 \neq 0$

What we Cannot Test with Either an F or a t-test

- We cannot test using this framework hypotheses which are not linear or which are multiplicative, e.g.

$$H_0 : \beta_2\beta_3 = 2 \text{ or } H_0 : \beta_2^2 = 1$$

cannot be tested.

The Relationship between the t and the F-Distributions

- Any hypothesis which could be tested with a t-test could have been tested using an F-test, but not the other way around.
- For example, consider the hypothesis

$$H_0 : \beta_2 = 0.5$$
$$H_1 : \beta_2 \neq 0.5$$

  We could have tested this using the usual t-test:

  $test\ stat = \frac{\hat{\beta}_2 - 0.5}{SE(\hat{\beta}_2)}$

  or it could be tested in the framework above for the F-test.
- Note that the two tests always give the same result since the t-distribution is just a special case of the F-distribution.
- For example, if we have some random variable Z, and $Z \sim t(T-k)$ then also $Z^2 \sim F(1, T-k)$

F-test Example

- Question: Suppose a researcher wants to test whether the returns on a company stock (y) show unit sensitivity to two factors (factor x_2 and factor x_3) among three considered. The regression is carried out on 144 monthly observations. The regression is $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$
  - What are the restricted and unrestricted regressions?
  - If the two RSS are 436.1 and 397.2 respectively, perform the test.
- Solution:

  Unit sensitivity implies H_0:$\beta_2 = 1$ and $\beta_3 = 1$. The unrestricted regression is the one in the question. The restricted regression is $(y_t - x_{2t} - x_{3t}) = \beta_1 + \beta_4 x_{4t} + u_t$ or letting $z_t = y_t - x_{2t} - x_{3t}$, the restricted regression is $z_t = \beta_1 + \beta_4 x_{4t} + u_t$

F-test Example (Cont'd)

In the F-test formula, T=144, k=4, m=2, $RRSS$=436.1, $URSS$=397.2

F-test statistic = 6.68. Critical value is an $F(2,140)$ = 3.07 (5%) and 4.79 (1%).

Conclusion: Reject H_0.

Data Mining

- Data mining is searching many series for statistical relationships without theoretical justification.
- For example, suppose we generate one dependent variable and twenty explanatory variables completely randomly and independently of each other.
- If we regress the dependent variable separately on each independent variable, on average one slope coefficient will be significant at 5%.
- If data mining occurs, the true significance level will be greater than the nominal significance level.

Goodness of Fit Statistics

- We would like some measure of how well our regression model actually fits the data.
- We have goodness of fit statistics to test this: i.e. how well the sample regression function (srf) fits the data.
- The most common goodness of fit statistic is known as R^2. One way to define R^2 is to say that it is the square of the correlation coefficient between y and $\hat{y}$.
- For another explanation, recall that what we are interested in doing is explaining the variability of y about its mean value, , i.e. the total sum of squares, TSS:

$$TSS = \sum_t (y_t - \bar{y})^2$$

- We can split the TSS into two parts, the part which we have explained (known as the explained sum of squares, ESS) and the part which we did not explain using the model (the RSS).

Defining R^2

- That is,

$$\begin{aligned} TSS &= ESS + RSS \\ \sum_t (y_t - \bar{y})^2 &= \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t \hat{u}_t^2 \end{aligned}$$

- Our goodness of fit statistic is

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

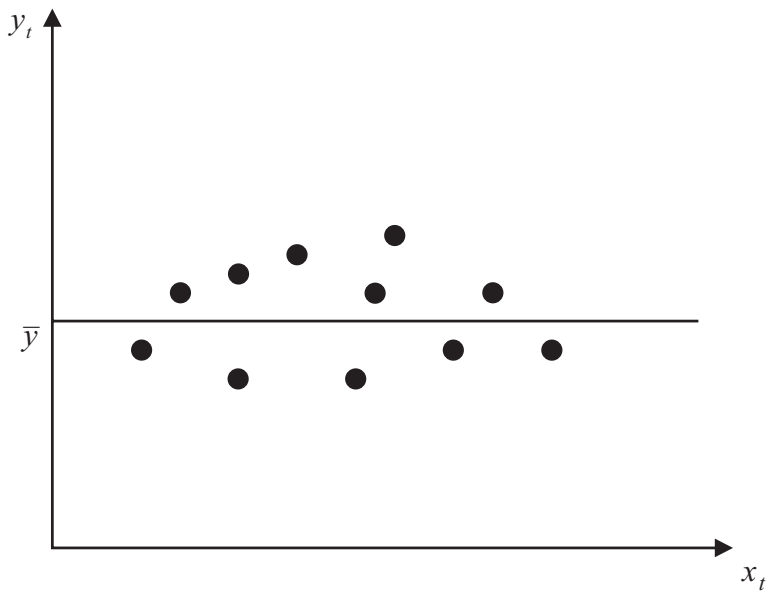
- But since $TSS = ESS + RSS$, we can also write

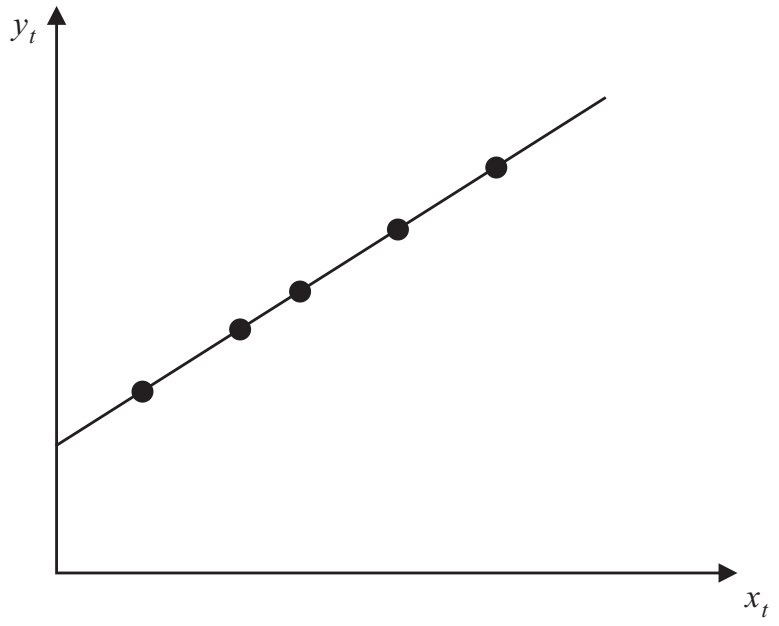
$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

- R^2 must always lie between zero and one. To understand this, consider two extremes

$$RSS = TSS \quad \text{i.e.} \quad ESS = 0 \quad \text{so} \quad R^2 = ESS/TSS = 0$$
$$ESS = TSS \quad \text{i.e.} \quad RSS = 0 \quad \text{so} \quad R^2 = ESS/TSS = 1$$

The Limit Cases: $R^2 = 0$ and $R^2 = 1$





Problems with R^2 as a Goodness of Fit Measure

- There are a number of them:

1. R^2 is defined in terms of variation about the mean of y so that if a model is reparameterised (rearranged) and the dependent variable changes, R^2 will change.
2. R^2 never falls if more regressors are added. to the regression, e.g. consider:

$$\begin{aligned} \textit{Regression } 1: y_t &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \\ \textit{Regression } 2: y_t &= \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \end{aligned}$$

   R^2 will always be at least as high for regression 2 relative to regression 1.

3. R^2 quite often takes on values of 0.9 or higher for time series regressions.

Adjusted R^2

- In order to get around these problems, a modification is often made which takes into account the loss of degrees of freedom associated with adding extra variables. This is known as $\bar{R}^2$, or adjusted $R^{2:}$

$$\bar{R}^2 = 1 - \left[\frac{T-1}{T-k}(1-R^2)\right]$$

- So if we add an extra regressor, k increases and unless R^2 increases by a more than offsetting amount, $\bar{R}^2$ will actually fall.

- There are still problems with the criterion:

  1. A "soft" rule
  2. No distribution for $\bar{R}^2$ or R^2

A Regression Example: Hedonic House Pricing Models

- Hedonic models are used to value real assets, especially housing, and view the asset as representing a bundle of characteristics.
- Des Rosiers and Thérialt (1996) consider the effect of various amenities on rental values for buildings and apartments 5 sub-markets in the Quebec area of Canada.
- The rental value in Canadian Dollars per month (the dependent variable) is a function of 9 to 14 variables (depending on the area under consideration). The paper employs 1990 data, and for the Quebec City region, there are 13,378 observations, and the 12 explanatory variables are:

Hedonic House Pricing Models: Variable Definitions

LnAGE	log of the apparent age of the property
NBROOMS	number of bedrooms
AREABYRM	area per room (in square metres)
ELEVATOR	a dummy variable = 1 if the building has an elevator; 0 otherwise
BASEMENT	a dummy variable = 1 if the unit is located in a basement; 0 otherwise
OUTPARK	number of outdoor parking spaces
INDPARK	number of indoor parking spaces
NOLEASE	a dummy variable = 1 if the unit has no lease attached to it; 0 otherwise
LnDISTCBD	log of the distance in kilometres to the central business district (CBD)

Hedonic House Pricing Models: Variable Definitions (Cont'd)

SINGLPAR	percentage of single parent families in the area where the building stands
DSHOPCNTR	distance in kilometres to the nearest shopping centre
VACDIFF1	vacancy difference between the building and the census figure

- The coefficient estimates themselves show the Canadian dollar rental price per month of each feature of the dwelling.

Hedonic House Price Results Dependent Variable: Canadian Dollars per Month

Variable	Coefficient	t-ratio	A priori sign expected
Intercept	282.21	56.09	+
LnAGE	−53.10	−59.71	−
NBROOMS	48.47	104.81	+
AREABYRM	3.97	29.99	+
ELEVATOR	88.51	45.04	+
BASEMENT	−15.90	−11.32	−
OUTPARK	7.17	7.07	+
INDPARK	73.76	31.25	+
NOLEASE	−16.99	−7.62	−
LnDISTCBD	5.84	4.60	−
SINGLPAR	−4.27	−38.88	−
DSHOPCNTR	−10.04	−5.97	−
VACDIFF1	0.29	5.98	−

Notes: Adjusted $R^2 = 0.651$; regression F-statistic = 2082.27.

Source: Des Rosiers and Thérialt (1996). Reprinted with permission of American Real Estate Society.

Tests of Non-nested Hypotheses

- All of the hypothesis tests concluded thus far have been in the context of "nested" models.
- But what if we wanted to compare between the following models?

$$\begin{aligned} \text{Model 1: } y_t &= \alpha_1 + \alpha_2 x_2 t + u_t \\ \text{Model 2: } y_t &= \beta_1 + \beta_2 x_{3t} + v_t \end{aligned}$$

- We could use R^2 or adjusted R^2, but what if the number of explanatory variables were different across the 2 models?
- An alternative approach is an encompassing test, based on examination of the hybrid model:

$$\text{Model 3: } y_t = \gamma_1 + \gamma_2 x_{2t} + \gamma_3 x_{3t} + w_t$$

Tests of Non-nested Hypotheses (Cont'd)

- There are 4 possible outcomes when Model 3 is estimated:
  - γ_2 is significant but γ_3 is not
  - γ_3 is significant but γ_2 is not
  - γ_2 and γ_3 are both statistically significant
  - Neither γ_2 nor γ_3 are significant
- Problems with encompassing approach
  - Hybrid model may be meaningless
  - Possible high correlation between x_2 and x_3.

Quantile Regression - Background

- Standard regression approaches effectively model the (conditional) mean of the dependent variable
- We could calculate from the fitted regression line the value that y would take for any values of the explanatory variables
- But this would be an extrapolation of the behaviour of the relationship between y and x at the mean to the remainder of the data
- This approach will often be suboptimal
- For example, there might be a non-linear (e.g., $\cap$-shaped) relationship between x and y
- Estimating a standard linear regression model may lead to seriously misleading estimates of this relationship as it will ́91average' the positive and negative effects.

Quantile Regression – Background 2

- It would be possible to include non-linear (i.e. polynomial) terms in the regression model (for example, squared, cubic, . . . terms)
- But quantile regressions represent a more natural and flexible way to capture the complexities by estimating models for the conditional quantile functions
- Quantile regressions can be conducted in both time-series and cross-sectional contexts
- It is usually assumed that the dependent variable, often called the response variable, is independently distributed and homoscedastic
- Quantile regressions are more robust to outliers and non-normality than OLS regressions

Quantile Regression – Background 3

- Quantile regression is a non-parametric technique since no distributional assumptions are required to optimally estimate the parameters
- The notation and approaches commonly used in quantile regression modelling are different to those that we are familiar with in financial econometrics
- Increased interest in modelling the 'tail behaviour' of series have spurred applications of quantile regression in finance
- A common use of the technique here is to value at risk modelling
- This seems natural given that the models are based on estimating the quantile of a distribution of possible losses.

Quantiles – A Definition

- Quantiles, denoted τ , refer to the position where an observation falls within an ordered series for y , for example:
  - The median is the observation in the very middle
  - The (lower) tenth percentile is the value that places 10% of observations below it (and therefore 90% of observations above)
- More precisely, we can define the τ-th quantile, $Q(\tau)$, of a random variable y having cumulative distribution $F(y)$ as

$$Q(\tau) = \inf y : F(y) \geq \tau$$

  where inf refers to the infimum, or the 'greatest lower bound', which is the smallest value of y satisfying the inequality
- By definition, quantiles must lie between zero and one
- Quantile regressions effectively model the entire conditional distribution of y given the explanatory variables.

Estimation of Quantile Functions

- The OLS estimator finds the mean value that minimises the RSS and minimising the sum of the absolute values of the residuals will yield the median
- The absolute value function is symmetrical so that the median always has the same number of data points above it as below it
- If the absolute residuals are weighted differently depending on whether they are positive or negative, we can calculate the quantiles of the distribution
- To estimate the τ-th quantile, we would set the weight on positive observations to τ and that on negative observations to 1-τ
- We can select the quantiles of interest and common choices would be 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95
- The fit is not good for values of τ too close to its limits of 0 and 1.

Estimation of Quantile Functions 2

- We could write the minimisation problem for a set of quantile regression parameters $\hat{\beta}_\tau$, each element of which is a $k \times 1$ vector, as

$$\hat{\beta}_\tau = \arg\min_{\beta} \left(\sum_{i:y_i > \beta x_i} \tau |y_i - \beta x_i| + \sum_{i:y_i < \beta x_i} (1-\tau)|y_i - \beta x_i| \right)$$

- As above, for the median, $\tau = 0.5$ and the weights are symmetric but for all other quantiles they will be asymmetric
- This optimisation problem can be solved using a linear programming representation via the simplex algorithm or within the generalised method of moments framework.

Quantile Regression – How not to do it

- As an alternative to quantile regression, it would be tempting to think of partitioning the data and running separate regressions on each of them
  - For example, dropping the top 90% of the observations on y and the corresponding data points for the xs, and running a regression on the remainder
- However, this process, tantamount to truncating the dependent variable, would be wholly inappropriate
  - It could lead to potentially severe sample selection biases
- In fact, quantile regression does not partition the data
  - All observations are used in the estimation of the parameters for every quantile

Quantile Regression Example

- A study by Bassett and Chen (2001) performs a style attribution analysis for a mutual fund and, for comparison, the S&P500 index
- To examine how a portfolio's exposure to various styles varies with performance, they use a quantile regression approach
- They conduct a style analysis by regressing the returns of a fund on the returns of a large growth portfolio, the returns of a large value portfolio, the returns of a small growth portfolio, and the returns of a small value portfolio
- These style portfolio returns are based on the Russell style indices
- The parameter estimates on each of these style-mimicking portfolio returns will measure the extent to which the fund is exposed to that style.

Quantile Regression Example – Discussion of Results

- We can determine the actual investment style of a fund without knowing anything about its holdings purely based on an analysis of its returns ex post and their relationships with the returns of style indices
- The results are shown from a standard OLS regression and quintile regressions for $\tau = 0.1$, 0.3, 0.5 (i.e. the median), 0.7, and 0.9
- The data are observed over the five years to December 1997 with standard errors based on a bootstrapping procedure
- Notice that the sum of the style parameters for a given regression is always one (except for rounding errors)
- The OLS results (column 2) show that the mean return has by far its biggest exposure to large value stocks (and this parameter estimate is also statistically significant).

Quantile Regression Example – Discussion of Results 2

- Comparing the mean (OLS) results with those for the median, $Q(0.5)$, the latter show much higher exposure to large value, less to small growth and none at all to large growth.
- We can examine the factor tilts as we move through the quantiles from left ($Q(0.1)$) to right ($Q(0.9)$)
  - The loading on large growth monotonically falls from 0.31 at $Q(0.1)$ to 0.01 at $Q(0.9)$ while the loadings on large value and small growth substantially increase
  - The loading on small value falls from 0.31 at $Q(0.1)$ to -0.51 at $Q(0.9)$
  - It is obvious that the intercept (coefficient on the constant) estimates should be monotonically increasing from left to right since the quantile regression effectively sorts on average performance
  - The intercept can be interpreted as the performance expected if the fund had zero exposure to all of the styles.

Quantile Regression Example – Table of Results

OLS and quantile regression results for the Magellan fund

	OLS	$Q(0.1)$	$Q(0.3)$	$Q(0.5)$	$Q(0.7)$	$Q(0.9)$
Large growth	0.14	0.35	0.19	0.01	0.12	0.01
	(0.15)	(0.31)	(0.22)	(0.16)	(0.20)	(0.22)
Large value	0.69	0.31	0.75	0.83	0.85	0.82
	(0.20)	(0.38)	(0.30)	(0.25)	(0.30)	(0.36)
Small Growth	0.21	−0.01	0.10	0.14	0.27	0.53
	(0.11)	(0.15)	(0.16)	(0.17)	(0.17)	(0.15)
Small Value	−0.03	0.31	0.08	0.07	−0.31	−0.51
	(0.20)	(0.31)	(0.27)	(0.29)	(0.32)	(0.35)
Constant	−0.05	−1.90	−1.11	−0.30	0.89	2.31
	(0.25)	(0.39)	(0.27)	(0.38)	(0.40)	(0.57)

Notes: Standard errors in parentheses. Source: Bassett and Chen (2001). Reprinted with the permission of Springer-Verlag.

Factor Models and Principal Components Analysis

- Factor models are employed as dimensionality reduction techniques in situations where we have a large number of closely related variables
- They decompose the structure of a set of series into factors that are common and a proportion that is specific to each series (idiosyncratic)
- There are two types of such models: economic and mathematical factor models
- The key distinction between the two is that the factors are observable for the former but are latent (unobservable) for the latter

Factor Models and Principal Components Analysis (Cont'd)

- Observable factor models include the APT model of Ross (1976)
- The most common mathematical model is principal components analysis
- PCA may be useful where explanatory variables are closely related – for example, in the context of near multicollinearity.

How PCA Works

- If there are k explanatory variables in the regression model, PCA will transform them into k uncorrelated new variables
- Suppose that the original explanatory variables are denoted x_1, x_2,...,x_k, and denote the principal components by p_1, p_2,..., p_k
- These principal components are independent linear combinations of the original data:

$$\begin{aligned} p_1 &= \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1k}x_k \\ p_2 &= \alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2k}x_k \\ \cdots &\quad \cdots \quad \cdots \quad \cdots \\ p_k &= \alpha_{k1}x_1 + \alpha_{k2}x_2 + \cdots + \alpha_{kk}x_k \end{aligned}$$

  where α_{ij} are coefficients to be calculated, representing the coefficient on the j^{th} explanatory variable in the i^{th} principal component.
- These coefficients are factor loadings.

PCA – More Details

- The sum of the squares of the coefficients for each component will be one
- Constructing the components is a purely mathematical exercise in constrained optimisation, and thus no assumption is made concerning the structure, distribution, or other properties of the variables
- The principal components are derived in such a way that they are in descending order of importance.
- Although there are k principal components, if there is some collinearity between the original explanatory variables, it is likely that some of the principal components will account for so little of the variation that they can be discarded.

Principal Components as Eigenvalues

- The principal components can also be understood as the eigenvalues of $(X'X)$, where X is the matrix of observations on the original variables
- If the ordered eigenvalues are denoted $\lambda_i (i = 1, \ldots, k)$, the ratio:

$$\phi_i = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i}$$

  gives the proportion of the total variation in the original data explained by the principal component i
- If only the first r$(0 < r < k)$ principal components are useful in explaining the variation of $(X'X)$ and are retained, the remaining k-r components would be discarded.

Principal Components as Eigenvalues

- The regression finally estimated, after the principal components have been formed, would be one of y on the first r principal components:

$$y_t = \gamma_0 + \gamma_1 p_{1t} + ... + \gamma_r p_{rt} + u_t$$

- In this way, the principal components are argued to keep most of the important information contained in the original explanatory variables, but are orthogonal
- The principal component estimates from this regression will be biased, although they will be more efficient than the OLS ones since redundant information has been removed
- The principal component coefficient estimates will simply be linear combinations of the original OLS estimates.

PCA Example: An Application to Interest Rates

- Researchers may wish to include interest rates on a large number of different assets in order to reflect the variety of investment opportunities open to investors
- However, market interest rates are likely to be highly correlated
- One approach would be to use PCA on several related interest rate series to determine whether they are actually closely related or not
- Fase (1973) conducted a study of monthly Dutch market interest rates from January 1962 until December 1970 (108 months)
- The money market instruments investigated were:
  - Call money, 3-month Treasury paper, 1-year T-paper, 2-year T-paper, 3-year T-paper, 5-year T-paper, 3-month loans to local authorities, 1-year loans to local authorities, Eurodollar deposits, Netherlands Bank official discount rate.

PCA Example: The Principal Components

- Prior to analysis, each series was standardised to have zero mean and unit variance
- The three largest of the ten eigenvalues are given in the following table
- The first principal component is sufficient to describe the common variation in these Dutch interest rates
- The 1^{st} component is able to explain over 90% of the variation for all samples

	Monthly data			*Quarterly data*
	Jan 62–Dec 70	Jan 62–Jun 66	Jul 66–Dec 70	Jan 62–Dec 70
λ_1	9.57	9.31	9.32	9.67
λ_2	0.20	0.31	0.40	0.16
λ_3	0.09	0.20	0.17	0.07
ϕ_1	95.7%	93.1%	93.2%	96.7%

Source: Fase (1973). Reprinted with the permission of Elsevier.

PCA Example: The Factor Loadings

- The factor loadings (coefficient estimates) for the first two ordered components are given in the table below
- The loadings on each factor making up the first principal component are all positive
- Since each series has been standardised, the coefficients α_{j1} and α_{j2} can be interpreted as the correlations between the interest rate j and the first and second principal components, respectively
- The factor loadings for each interest rate series on the first component are all very close to one
- Fase (1973) therefore argues that the first component can be interpreted simply as an equally weighted combination of all of the market interest rates.

PCA Example: The Factor Loadings 2

- The second component, which explains much less of the variability of the rates, shows a factor loading pattern of positive coefficients for the Treasury paper series and negative or almost zero values for the other series
- Fase (1973) argues that this is owing to the characteristics of the Dutch Treasury instruments that they rarely change hands and have low transactions costs, and therefore have less sensitivity to general interest rate movements
- Also, they are not subject to default risks in the same way as, for example, Eurodollar deposits
- Therefore, the second principal component is broadly interpreted as relating to default risk and transactions costs.

PCA Example: The Factor Loadings Presented

j	Debt instrument	α_{j1}	α_{j2}
1	Call money	0.95	−0.22
2	3-month Treasury paper	0.98	0.12
3	1-year Treasury paper	0.99	0.15
4	2-year Treasury paper	0.99	0.13
5	3-year Treasury paper	0.99	0.11
6	5-year Treasury paper	0.99	0.09
7	Loans to local authorities: 3-month	0.99	−0.08
8	Loans to local authorities: 1-year	0.99	−0.04
9	Eurodollar deposits	0.96	−0.26
10	Netherlands Bank official discount rate	0.96	−0.03
	Eigenvalue, λ_i	9.57	0.20
	Proportion of variability explained by eigenvalue i, ϕ_i(%)	95.7	2.0

Source: Fase (1973). Reprinted with the permission of Elsevier.

Limitations of PCA

- A change in the units of measurement of x will change the principal components
- It is thus usual to transform all of the variables to have zero mean and unit variance prior to applying PCA
- The principal components usually have no theoretical motivation or interpretation whatsoever
- The r principal components retained from the original k are the ones that explain most of the variation in x, but these components might not be the most useful as explanations for y.