8. Time Series and Prediction

Definition: A time series is given by a sequence of the values of a variable observed at sequential points in time.

e.g. daily maximum temperature, end of day share prices, annual GDP.

For convenience, it will be assumed that the time between observations is constant,

e.g. the unit of time when giving end of day share prices is a working day. However, it should be noted that it is possible that changes in share prices show more volatility (variation) between closing on Friday and closing on Monday than between Monday and Tuesday.

Time Series and Prediction

It is clear from the previous chapter that regression can be used to analyse a time series (see Example 7.3).

However, time series analysis allows us to carry out more sophisticated analysis.

The first question to ask when analysing a time series is "is there a trend in the data and if so what sort of trend is there?"

8.1 Trends

The following cases are commonly met

1. No trend, i.e. variation around a mean level.
2. A linear trend.
3. An exponential trend.
4. A seasonal trend (seasonal variation).

Trends of type 2 and 3 are called long-term trends.

Trends

By looking at a scatter plot of the data (with time on the x-axis), we can decide what sort of trend(s) is/are present.

For an example of a linear trend look at the relationship between time and ln(GDP) in Example 7.3.

For an example of an exponential trend look at the relationship between time and GDP in Example 7.3.

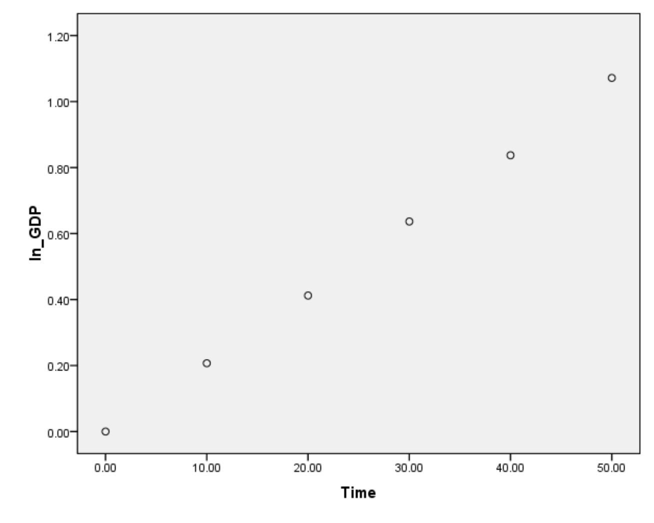
One obvious variable that shows seasonal variation is the daily maximum temperature. When seasonal variation occurs, it is important to determine the length of a cycle (daily high temperature shows an annual cycle, i.e. approximately 365 days or 12 months).

Trends

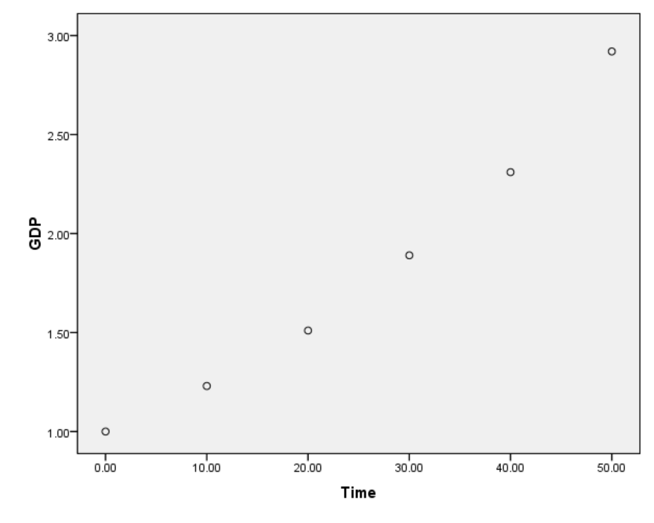
Some variables may show a combination of trends. For example, the number of air passengers has been growing exponentially since the 1950s. In addition, there are seasonal peaks in the summer and at christmas.

However, such a combination of trends is beyond the scope of this module.

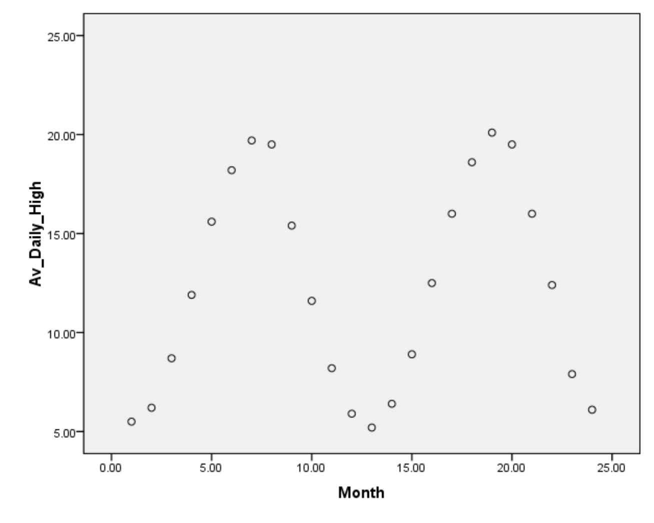
Linear Trend



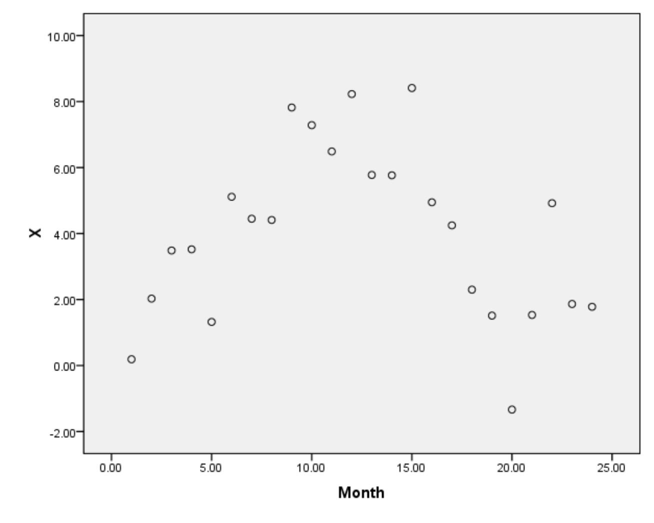
Exponential Trend



Seasonal Variation



Time Series with no Trend



Time Series without a Trend

Such a series varies around some mean.

The simplest means of predicting future observations is to assume that future observations will be equal to the mean of the observed values.

This is the most sensible estimate when it can be assumed that the observations in the time series are i.i.d. (independent observations from the same distribution).

However, in many cases this will not be the case and we can improve our predictions.

Time Series without a Trend

For example, suppose the temperature at the equator shows no seasonal variation, but hot days are likely to followed by hot days and cold days are likely to be followed by cold days.

In this case, we may use a "moving average" approach to estimating future temperatures.

This means that e.g. the estimate of the daily high temperature for day $t+1$ based on the daily highs for days $1, 2, \ldots t$ and a moving average procedure with lag k is simply the average of last k daily highs observed.

8.2 Prediction and Smoothing Using Moving Averages

Suppose a moving average of order k is used, this will be denoted MA(k).

Note that the choice of the method to be used is outside the scope of this module. We consider only the derivations of predictions given the method accepted.

Suppose we have observations $X_1, X_2, \ldots, X_T$ where the subscript denotes time. It will be assumed that T is relatively large.

The predicted value of X_{T+k} made at time T is denoted by $\hat{X}_{T+k;T}$.

Smoothing using Moving Averages

The predicted value of X_{T+1} made at time T, $\hat{X}_{T+1;T}$ is simply the average of the last k observations

$$\hat{X}_{T+1;T} = \frac{X_{T-k+1} + X_{T-k+2} + \ldots + X_T}{k}.$$

Note that if we want to predict two steps into the future, we do not know what the observation X_{T+1} is, but we can replace it by the prediction made at time T, i.e.

$$\hat{X}_{T+2;T} = \frac{X_{T-k+2} + X_{T-k+3} + \ldots + X_T + \hat{X}_{T+1;T}}{k}.$$

Note that we are still averaging k terms.

Smoothing using Moving Averages

Similarly, we can predict three steps into the future by replacing the observations at times $T+1$ and $T+2$ by their predictions

$$\hat{X}_{T+3;T} = \frac{X_{T-k+3} + X_{T-k+4} + \ldots + X_T + \hat{X}_{T+1;T} + \hat{X}_{T+2;T}}{k}.$$

Predictions more than 3 steps into the future can be made in an analogous manner.

Note that in order to make a prediction i steps into the future, we must first make predictions $1, 2, \ldots, i-1$ steps into the future.

Smoothing using Moving Averages

Also, once X_{T+1} has been observed, the predictions of X_{T+2} and X_{T+3} can be updated by replacing the prediction of $\hat{X}_{T+1;T}$ by the actual value, i.e.

$$\hat{X}_{T+2;T+1} = \frac{X_{T-k+2} + X_{T-k+3} + \ldots + X_T + X_{T+1}}{k}$$

$$\hat{X}_{T+3;T+1} = \frac{X_{T-k+3} + X_{T-k+4} + \ldots + X_T + X_{T+1} + \hat{X}_{T+2;T}}{k}.$$

Obviously, the further we try to predict into the future, the less accurate our predictions.

Example 8.1

Suppose we have the following time series

t	1	2	3	4	5	6	7
X_t	2.02	1.86	2.68	2.59	2.43	2.26	1.68

Assuming there is no long-term trend in the series

i) Using a moving average of length 3, predict X_8 and X_9 based on these observations.

ii) Suppose X_8 is then observed to be 1.50. Update your estimate of X_9.

Example 8.1

i) Our prediction of X_8 made after observing X_7, denoted $\hat{X}_{8;7}$, is based on the average of the previous 3 observations.

$$\hat{X}_{8;7} = \frac{X_5 + X_6 + X_7}{3} = \frac{2.43 + 2.26 + 1.68}{3} \approx 2.12.$$

Example 8.1

Now we can predict X_9 using the prediction of X_8 and the previous 2 observations (thus averaging over 3 observations)

$$\hat{X}_{9;7} = \frac{X_6 + X_7 + \hat{X}_{8;7}}{3} = \frac{2.26 + 1.68 + 2.12}{3} \approx 2.02.$$

Example 8.1

iii) Once we have observed X_8, using the moving average approach, we can now update our prediction of X_9 by replacing the prediction $\hat{X}_{8;7}$ by the value of the observation itself. Our prediction of X_9 is now the mean of the previous three observations

$$\hat{X}_{9;8} = \frac{X_6 + X_7 + X_8}{3} = \frac{2.26 + 1.68 + 1.50}{3} \approx 1.81.$$

It should be noted that since X_8 is less than the predicted value $\hat{X}_{8;7}$, the updated prediction of X_9 is lower than the initial prediction (as would be expected).

8.3 Dealing with Seasonal Variation

Many time series show seasonal variation (cyclical variation).

Temperature shows cyclical variation over a day and (at larger scale) over a year.

It is thought that the economy shows some degree of cyclical behaviour (for example, government spending might show 5-year cycles if governments try to win elections by spending relatively more in pre-election years than in post-election years).

It is assumed in this section that there are no long-term trends in the data.

Seasonal Variation

If seasonal variation can be seen from a scatter plot, the first thing to do is establish the cycle length. In many cases this will be obvious, e.g. the period of daily high temperatures and the sales of toys will have a cycle length of one year.

In this case the time series will vary around some oscillating level describing the expected value at a certain time in the season.

We consider two models for the variation around the expected value: a) the level of variation is constant - the so called additive model and b) the level of variation is proportional to the expected value - the so called multiplicative model.

If observation can take negative values or the seasonal variation is not extreme, then one can always use the additive model. If there are sharp peaks in the seasonal variation, then the multiplicative model may well be more appropriate.

Seasonal Variation

We denote the length of the cycle by d. For example, if cycles are annual and we have monthly data, then $d = 12$. If we have daily data, then we may assume $d = 365$ (unless the time series covers a very long period, i.e. centuries).

It is assumed that our data correspond to an integer number k cycles, e.g. if we are dealing with annual cycles, then we have data for an integer number of years, i.e. the number of observations is kd, where k is an integer.

Define $X_{i,j}$ to be the observation in period j of cycle i. Thus we have observations $X_{1,1}, X_{1,2}, \ldots, X_{k,d}$.

Seasonal Variation

After that we calculate the mean of the observations taken in period j of the cycle, denoted $\overline{X}_j$, $j = 1, 2, \ldots d$.

For example, if we are dealing with monthly data and annual cycles, then $\overline{X}_1$ will be the mean of all the January observations, $\overline{X}_2$ will be the mean of all the February observations and so on.

Seasonal Variation - Additive Model

The model for the observation made in period j of cycle i is given by

$$X_{i,j} = \overline{X}_j + \epsilon_{i,j},$$

where $\epsilon_{i,j}$ is the random deviation in period j of cycle i (expected value 0).

Predictions under an Additive Model of Seasonal Variation

In order to make predictions, we need first to be able to define a sequence of "deviations" which have no long-term or seasonal trend.

Suppose we have obtained a model of the form

$$X_{i,j} = \overline{X}_j + \epsilon_{i,j}.$$

This can be rearranged into the form

$$\epsilon_{i,j} = X_{i,j} - \overline{X}_j.$$

Since the ϵ are random deviations, if there is only seasonal variation, then the sequence of ϵ will have no trend.

Predictions under an Additive Model of Seasonal Variation

When we assume that these deviations are independent, then we may predict any observation made in period j of the cycle using

$$\hat{X}_{i,j} = \overline{X}_j.$$

i.e. the prediction is equal to the mean for that period of the season. Note this prediction is independent of the time at which it is made.

However, as with time series with no trend, neighbouring deviations may be correlated e.g. after a very hot June, one might expect a hot July.

If this is expected, then we can predict future residuals using a moving average procedure. This can then be used to predict the appropriate value of the observation.

Example 8.2

Suppose we have the following quarterly data on i-tunes sales (in billions of $).

Q	1	2	3	4	5	6	7	8	9	10	11	12
S	1.3	1.5	1.1	2.2	1.4	1.4	1.0	2.0	1.5	1.3	1.2	2.4

i) Derive an additive model of seasonal variation for the sales made in each quarter of the year.

ii) Assuming the deviations to the model are all independent, use this model to estimate sales in each quarter of the following year

iii) Using a moving average of order 2 to predict residuals, estimate sales in the first two quarters of the next year.

Example 8.2

i) Since we have quarterly data $d = 4$ and we have data for 3 years. Firstly, we calculate the mean volume for each quarter. Thus

$$\begin{aligned}
\overline{X}_1 &= \frac{X_{1,1} + X_{2,1} + X_{3,1}}{3} = 1.4 \\
\overline{X}_2 &= \frac{X_{1,2} + X_{2,2} + X_{3,2}}{3} = 1.4 \\
\overline{X}_3 &= \frac{X_{1,3} + X_{2,3} + X_{3,3}}{3} = 1.1 \\
\overline{X}_4 &= \frac{X_{1,4} + X_{2,4} + X_{3,4}}{3} = 2.2,
\end{aligned}$$

where $X_{i,j}$ are the sales in quarter j of year i, i.e. $X_{2,1}$ will be the sales in quarter 5, $X_{3,1}$ the sales in quarter 9.

Example 8.2

Hence, the model is

$$
\begin{aligned}
X_{i,1} &= 1.4 + \epsilon_{i,1}, \quad \text{sales in Quarter 1} \\
X_{i,2} &= 1.4 + \epsilon_{i,2}, \quad \text{sales in Quarter 2} \\
X_{i,3} &= 1.1 + \epsilon_{i,3}, \quad \text{sales in Quarter 3} \\
X_{i,4} &= 2.2 + \epsilon_{i,4}, \quad \text{sales in Quarter 4}
\end{aligned}
$$

where i is some integer.

Example 8.2

ii) a) Based on this model, if the ϵ are independent, the best estimates of future sales are the mean values for the relevant quarter. Note the 13th quarter corresponds to the first quarter of year 4, the 14th quarter corresponds to the second quarter of year 4 etc.

$$\hat{X}_{4,1} = \overline{X}_1 = 1.4, \quad \hat{X}_{4,2} = \overline{X}_2 = 1.4,$$
$$\hat{X}_{4,3} = \overline{X}_3 = 1.1, \quad \hat{X}_{4,4} = \overline{X}_4 = 2.2.$$

Example 8.2

b) The residuals are given by

$$\epsilon_{i,j} = X_{i,j} - \overline{X}_j,$$

In particular, we need to calculate the final two residuals. We have

$$\begin{aligned}\epsilon_{3,3} &= X_{3,3} - \overline{X}_3 = 1.2 - 1.1 = 0.1 \\ \epsilon_{3,4} &= X_{3,4} - \overline{X}_4 = 2.4 - 2.2 = 0.2.\end{aligned}$$

Example 8.2

Using a moving average of order 2, it follows that the estimate of the residual for the first quarter of year 4 is

$$\hat{\epsilon}_{4,1} = \frac{\epsilon_{3,3} + \epsilon_{3,4}}{2} = \frac{0.1 + 0.2}{2} = 0.15.$$

The estimate of the residual for the second quarter of year 4 is

$$\hat{\epsilon}_{4,2} = \frac{\epsilon_{3,4} + \hat{\epsilon}_{4,1}}{2} = \frac{0.2 + 0.15}{2} = 0.175$$

Example 8.2

It follows that the estimates of the sales in the first two quarters of the next year are given by

$$
\begin{aligned}
\hat{X}_{4,1} &= \overline{X}_1 + \hat{\epsilon}_{4,1} = 1.4 + 0.15 = 1.55 \\
\hat{X}_{4,2} &= \overline{X}_2 + \hat{\epsilon}_{4,2} = 1.4 + 0.175 = 1.575
\end{aligned}
$$

A Multiplicative Model

Sales may exhibit greater variability in periods of high sales than in periods of low sales. In this case, we should use a multiplicative model.

Suppose observation $X_{i,j}$ is made in period j of cycle i, we have a model of the form

$$X_{i,j} = \overline{X}_j(1 + \epsilon_{i,j}), \tag{1}$$

where $\epsilon_{i,j}$ now represents the relative deviation at time t rather the absolute deviation as in the additive model. The expected value of $\epsilon_{i,j}$ is zero.

A Multiplicative Model

Assuming that the residuals are independent, predictions can be made in the same way as under the additive model

i.e. the prediction of an observation to be made in period j of a cycle is equal to the mean of the observations observed at that period in the cycle.

If the residuals from this model are correlated, then we can predict future residuals using a moving average procedure of order 2 and use these to make predictions of future observations.

Return to Example 8.2

In order to predict $X_{4,1}$ and $X_{4,2}$ using a moving average of order 2 to predict future residuals, we first need to calculate the final two residuals. The residuals are given by

$$\epsilon_{i,j} = \frac{X_{i,j}}{\overline{X}_j} - 1$$

Hence, we have

$$\begin{aligned}
\epsilon_{3,3} &= \frac{X_{3,3}}{\overline{X}_3} - 1 = \frac{1.2}{1.1} - 1 = 0.0909 \\
\epsilon_{3,4} &= \frac{X_{3,4}}{\overline{X}_4} - 1 = \frac{2.4}{2.2} - 1 = 0.0909.
\end{aligned}$$

Return to Example 8.2

It follows that the estimates of the residuals for the first two quarters of year 4 are given by

$$\hat{\epsilon}_{4,1} = \frac{\epsilon_{3,3} + \epsilon_{3,4}}{2} = 0.0909$$

$$\hat{\epsilon}_{4,2} = \frac{\epsilon_{3,4} + \hat{\epsilon}_{4,1}}{2} = 0.0909$$

Return to Example 8.2

It follows that the estimates of the sales for the first two quarters of year 4 are given by

$$\hat{X}_{4,1} = \overline{X}_1(1 + \hat{\epsilon}_{4,1}) = 1.4 \times 1.0909 = 1.527$$
$$\hat{X}_{4,2} = \overline{X}_2(1 + \hat{\epsilon}_{4,2}) = 1.4 \times 1.0909 = 1.527$$

8.4 Dealing with Linear Trends

If a scatter plot indicates that a trend is linear, then we may **initially** analyze the time series by calculating the regression line (using time as the dependent variable) .

The model obtained in this way is of the form

$$X_t = \beta_0 + \beta_1 t + \epsilon_t,$$

where ϵ_t is the residual at time t. Rearranging, we obtain

$$\epsilon_t = X_t - (\beta_0 + \beta_1 t).$$

Using a Moving Average Approach for Linear Trends

Now we define a moving average approach to predicting future observations.

Suppose the value of X_t can be modelled using

$$X_t = \beta_0 + \beta_1 t + \epsilon_t.$$

Define $Y_t = X_t - X_{t-1}$ to be the "slope" at time t (i.e. the increase in X between time $t-1$ and time t. It follows that

$$X_t - X_{t-1} = \beta_1 + \epsilon_t - \epsilon_{t-1}.$$

Since the ϵ are random variables with expected value 0, it follows that the difference between neighbouring observations of X vary around a constant level β_1, i.e. the series of Y shows no trend.

Linear Trends - Predicting Observations using a Moving Average

If a scatter plot indicates that a linear function will estimate the trend well, one obvious way of estimating X_{t+1} based on the observations $X_1, X_2, \ldots, X_t$ can be obtained by assuming that X increases by the same amount between time t and $t+1$ as it did between time $t-1$ and t, i.e.

$$\hat{X}_{t+1;t} = X_t + Y_t = X_t + (X_t - X_{t-1}) = 2X_t - X_{t-1}.$$

This idea can be also be used to predict observations further in the future, i.e the prediction for X_{t+k} made at time t is given by

$$\hat{X}_{t+k;t} = X_t + kY_t = X_t + k(X_t - X_{t-1}) = (k+1)X_t - kX_{t-1}.$$

Linear Trends - Predicting Observations using a Moving Average

If the Y are relatively variable, it will be better to smooth their values using a moving average of order k. The slope at future intervals can then be estimated using the appropriate moving average.

For example, if we choose a moving average of order k, our estimate of the slope at time t is given by U_t, where

$$U_t = \frac{Y_{t-k+1} + Y_{t-k+2} + \ldots + Y_t}{k} = \frac{X_t - X_{t-k}}{k},$$

i.e. this is simply the average slope over the past k intervals. Hence, our prediction of X_{t+1} is given by

$$\hat{X}_{t+1;t} = X_t + U_t$$

Iterating this we obtain

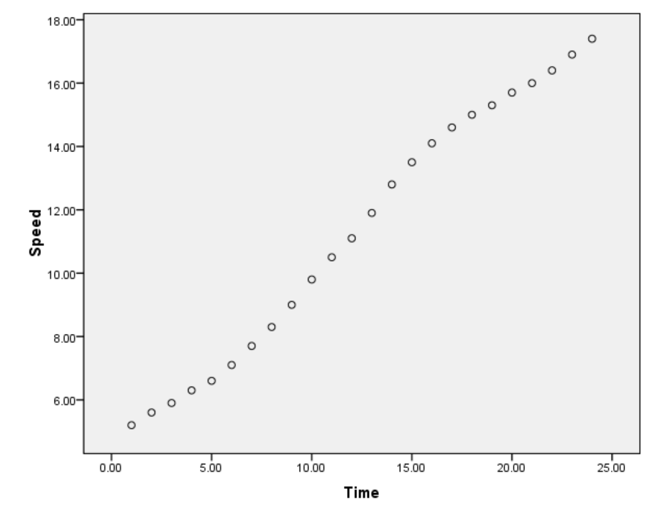
$$\hat{X}_{t+k;t} = X_t + kU_t$$

Example 8.3

The data in lintrend.sav give the speed of a ball rolling down a hill according to time. t goes from 1 to 24.

A plot of speed against time is on the next slide.

Example 8.3


18.00
16.00
14.00
12.00
10.00
8.00
6.00
Speed
0.00
5.00
10.00
15.00
20.00
25.00
Time

Example 8.3

The graph seems S shaped rather than linear.

However, the curvature of the plot is not great and a linear trend would fit relatively well (any model taking into account the curvature of the plot would have to be clearly more complex).

The regression line for speed as a function of time is given by (see next slide for the SPSS output)

$$Speed = 4.165 + 0.576t.$$

Example 8.3 - SPSS output describing the regression model

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4.165	.187		22.304	.000
	Time	.576	.013	.994	44.054	.000

a. Dependent Variable: Speed

Example 8.3

We might use a moving average procedure of order 4 to estimate the slope and hence future observations.

In order to do this, we need the final five observations (see next slide).

Example 8.3

Time	20	21	22	23	24
Speed (X_t)	15.7	16.0	16.4	16.9	17.4

Example 8.3

Using the regression model, the predictions for the speed at time 25 and 26 are given by

$$\hat{X}_{25} = 4.165 + 0.576 \times 25 = 18.565$$
$$\hat{X}_{26} = 4.165 + 0.576 \times 26 = 19.141$$

Looking at the final five observations, these predictions are clearly overestimates.

Example 8.3

The smoothed estimate of the slope is

$$U_{24} = \frac{X_{24} - X_{20}}{4} = \frac{17.4 - 15.7}{4} = 0.425.$$

It follows that the predictions of the next two observations are given by

$$\begin{aligned} \hat{X}_{25;24} &= X_{24} + U_{24} = 17.825 \\ \hat{X}_{26;24} &= X_{24} + 2U_{24} = 18.25 \end{aligned}$$

Comparison of Regression and Time Series Analysis

In general, the use of such a moving average procedure will be more accurate than the use of regression.

This is due to the fact that the moving average procedure predicts on the basis of extrapolating the present observation based on recent trends.

Using regression, we are extrapolating **an estimate of the present observation** on the basis of a long-term trend which may not be indicative of the current short-term trend.

Suppose the series X_t shows exponential growth.

Arguing as in the section on exponential regression, the series $\ln(X_t)$ will exhibit linear growth.

Hence, we use the techniques outlined in Section 8.4 to analyse and predict the series $\ln(X_t)$.

By exponentiating, we then obtain a model and predictions for the series X_t.

The following (real) data give estimates of Ireland's GDP per capita in thousands of Dollars (X_t) from 2000 (year 0) to 2009 (year 9).

i) Derive a regression model for X_t as an exponential function of t.

ii) Using this regression model, predict GDP per capita in 2010

iii) Using a moving average procedure of order 2, estimate GDP per capita in 2010.

Exponential Growth

Year	0	1	2	3	4
GDP	20.3	21.6	28.5	30.5	29.6
ln(GDP)	3.01	3.07	3.35	3.42	3.39

Year	5	6	7	8	9
GDP	31.9	41.1	44.5	46.6	45.1
ln(GDP)	3.46	3.72	3.80	3.84	3.81

Exponential Growth

Using SPSS, we obtain the following regression model for $V_t = \ln(X_t)$.

$$\ln(X_t) = V_t = 3.056 + 0.096t.$$

Exponentiating, we obtain

$$X_t = \exp(3.056 + 0.096t) = e^{3.056}e^{0.096t} = 21.24e^{0.096t}.$$

Exponential Growth

Using this to estimate GDP per capita in 2010 (i.e. $t = 10$), we obtain

$$\hat{X}_{10} = 21.24e^{0.96} = 55.48$$

Exponential Growth

Using a moving average of order 2, we calculate the estimate of the slope for the variable showing linear growth (here ln(GDP)) based on the final two intervals.

It follows that our estimate of the slope is

$$U_9 = \frac{X_9 - X_7}{2} = \frac{3.81 - 3.80}{2} = 0.005.$$

Exponential Growth

Hence, our estimate of V_{10} is

$$\hat{V}_{10;9} = 3.81 + 0.005 = 3.815.$$

Since $V_t = \ln(X_t) \Rightarrow \exp(V_t) = X_t$, it follows that our estimate of GDP per capita is

$$\hat{X}_{10;9} = e^{3.815} = 45.38.$$

This differs a lot from the estimate from the regression model.

Problems with using Regression Models to Predict

In general, using a moving average is a much more accurate means of predicting future values in a time series than extrapolating with the use of a regression model.

The moving average takes into account the fact that GDP in 2010 will be much more related to GDP in 2009 and the recent path of GDP.

A regression model predicts a time series on the basis of the whole series and does not take these local dependencies into account.