

Forecasting

AI Nosedal
University of Toronto

March 8, 2016

Forecasting Methods: An Overview

There are many forecasting methods available, these methods can generally be divided into three groups:

- 1. Judgemental methods.
- 2. Extrapolation (or Time Series) methods, and
- 3. Econometric (or causal) methods.

Extrapolation Methods

Extrapolation methods are quantitative methods that use past data of a time series variable—and nothing else, except possibly time itself—to forecast future values of the variable. The idea is that we can use past movements of a variable, such as a company sales to forecast its future values.

- The simplest time series model is the random model. A random model can be written as

$$Y(t) = \mu + \epsilon(t) \quad (1)$$

- Here, μ is a constant, the average of the $Y(t)$'s, and $\epsilon(t)$ is the residual (or error) term. We assume that the residuals have mean 0, variance σ^2 , and are probabilistically independent of one another.

Random Series (1)

The first is when the original time series is random. For example, when studying the time pattern of diameters of individual parts from a manufacturing process, we might discover that the successive diameters behave like a random time series.

Random Series (2)

The second situation where a random series occurs is more common. This is when we fit a model to a time series to obtain an equation of the form

$$Y(t) = \textit{fitted part} + \textit{residual}. \quad (2)$$

Although the fitted part varies from model to model, its essential feature is that it describes any underlying time series pattern in the original data. The residual is then whatever is left, and we hope that the series of residuals is random with mean $\mu = 0$.

Runs test

For each observation $Y(t)$ we associate a 1 if $Y(t) \geq \bar{y}$ and 0 if $Y(t) < \bar{y}$. A run is a consecutive sequence of 0's or 1's. Let T be the number of observations, let T_a above the mean, and let T_b the number below the mean. Also let R be the observed number of runs. Then it can be show that for a random series

$$E(R) = \frac{T + 2T_a T_b}{T} \quad (3)$$

$$Stdev(R) = \sqrt{\frac{2T_a T_b (2T_a T_b - T)}{T^2 (T - 1)}} \quad (4)$$

When $T > 20$, the distribution of R is roughly Normal.

Example: Runs test

Suppose that the successive observations are 87, 69, 53, 57, 94, 81, 44, 68, and 77, with mean $\bar{Y} = 70$. It is possible that this series is random. Does the runs test support this conjecture?

Example: Runs test (solution)

The preceding sequence has five runs: 1; 0 0 0; 1 1; 0 0; and 1. Then, we have $T = 9$, $T_a = 4$, $T_b = 5$ and $R = 5$. Under a randomness hypothesis,

$$E(R) = 5.44 \quad (5)$$

$$Stdev(R) = 1.38 \quad (6)$$

$$Z = \frac{R - E(R)}{Stdev(R)} = -0.32 \quad (7)$$

Exercise

Write a function in R that performs a run test. My version of this function will be posted on our website tomorrow.

```

runs<-function(y){
  T<-length(y);
  y.bar<-mean(y);
  y[y<y.bar]<-0;      # if y is less than y.bar then 0;
  y[y>=y.bar]<-1;    # if y is greater or equal than y.bar then 1;
  sign<-y;
  Ta<-sum(sign);
  Tb<-T-Ta;
  R<- sum(abs(sign[-T]-sign[-1]))+1;    #runs;
  E.R<-(T+2*Ta*Tb)/T;
  Stdev<-sqrt((2*Ta*Tb*(2*Ta*Tb-T ) )/( T^2*(T-1) ) );
  Z<- (R-E.R)/Stdev;
  P.value<-2*(1-pnorm(abs(Z)));

  list('T'=T, 'Ta'=Ta, 'Tb'=Tb, 'R'=R, 'Z'=Z,
       'E(R)'=E.R, 'P value'=P.value)
}

```

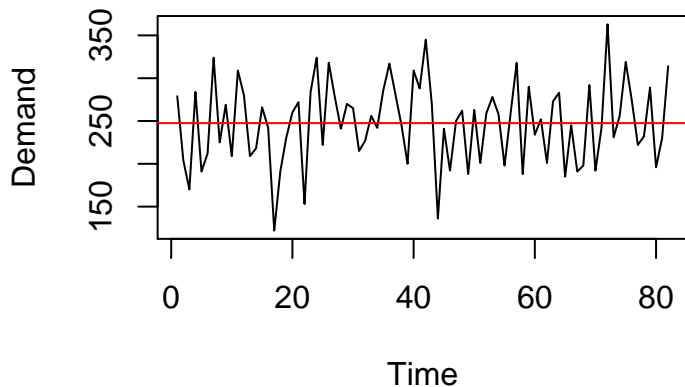
Example: Demand

The dollar demand for a certain class of parts at a local retail store has been recorded for 82 consecutive days. (See the file demand.txt) A time series plot of these demands appears in the next slide. The store manager wants to forecast future demands. In particular, he wants to know whether there is any significant time pattern to the historical demands or whether the series is essentially random.

Reading data

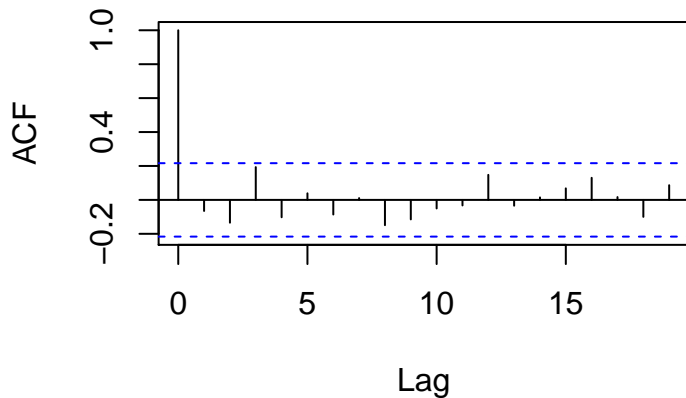
```
demand<-read.table(file="demand.txt",header=TRUE);  
  
names(demand);  
  
## [1] "Day"      "Demand"
```

Time Series Plot



A visual inspection of the time series graph in the previous slide shows that demands vary randomly around the sample mean of \$247.54 (shown as the horizontal centerline). The variance appears to be constant through time, and there are no obvious time series patterns. To check formally whether this apparent randomness holds, we perform the runs test and calculate the first 10 autocorrelations. (The associated correlogram appears in the next slide). The p-value for the runs test is relatively large 0.118 and none of the autocorrelations is significantly large. These findings are consistent with randomness.

Series demand\$Demand



Conclusion

For all practical purposes there is no time series pattern to these demand data. It is as if every day's demand is an independent draw from a distribution with mean \$247.54 and standard deviation \$47.78. Therefore, the manager might as well forecast that demand for any day in the future will be \$247.54. If he does so, about 95% of his forecasts should be within two standard deviations (about \$95) of the actual demands.

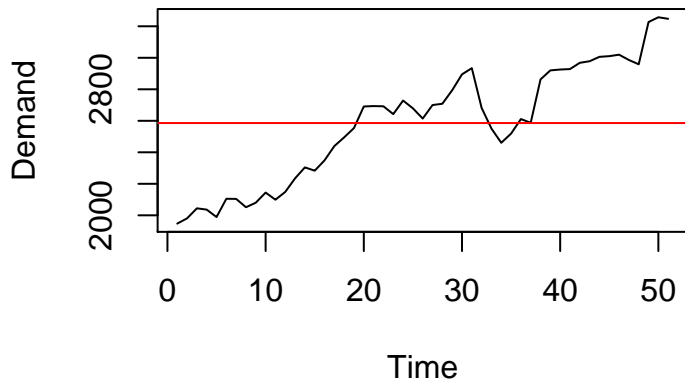
Random Walk Model

Random series are sometimes building blocks for other time series models. The model we now discuss, the random walk model, is an example of this. This type of behavior is typical of stock price data. For example, the graph in the next slide shows monthly Dow Jones averages from January 1988 through March 1992. (See the file dow.txt)

Reading data

```
dow<-read.table(file="dow.txt",header=TRUE);  
  
names(dow);  
  
## [1] "Date" "Dow"
```

Time Series Plot



More about Dow Jones

This series is not random, as can be seen from its gradual upward trend (Although the runs test and autocorrelations are not shown here, they confirm that the series is not random.)

If we were standing in March 1992 and were asked to forecast the Dow Jones average for the next few months, it is intuitive that we would not use the average of the historical values as our forecast. This forecast would probably be too low because the series has an upward trend. Instead, we would base our forecast on the most recent observation. This is exactly what the random walk model does.

Random walk (equation)

An equation for the random walk model is

$$Y_t = Y_{t-1} + \mu + \epsilon_t \quad (8)$$

where μ is a constant and ϵ_t is a random series with mean 0 and some standard deviation σ . If we let $DY_t = Y_t - Y_{t-1}$, the change in the series from time t to time $t - 1$, then we can write the random walk model as

$$DY_t = \mu + \epsilon_t \quad (9)$$

This implies that the differences form a random series with mean μ and standard deviation σ .

Random walk (cont.)

An estimate of μ is the average of the differences, labeled \bar{Y}_D , and an estimate of σ is the sample standard deviation of the differences, labeled s_D . In words, a series that behaves according to this random walk model has random differences, and the series tends to trend upward (if $\mu > 0$) or downward (if $\mu < 0$) by an amount μ each period. If we are standing in period t and want to make a forecast F_{t+1} of Y_{t+1} , then a reasonable forecast is

$$F_{t+1} = Y_t + \bar{Y}_D \quad (10)$$

That is, we add the estimated trend to the current observation to forecast the next observation.

Example: Dow Jones revisited

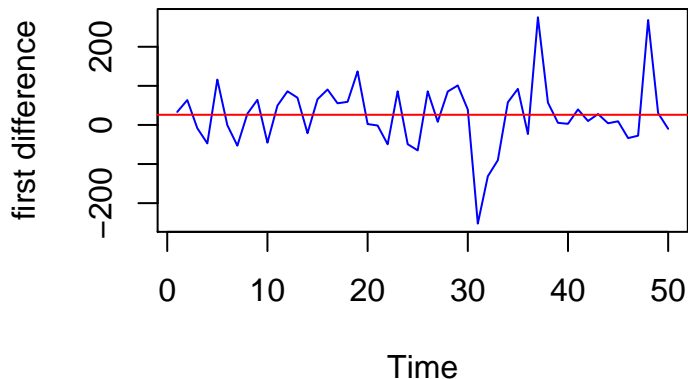
Given the monthly Dow Jones data in the file `dow.txt`, check that it satisfies the assumptions of a random walk, and use the random walk model to forecast the value for April 1992.

Solution

We have already seen that the Dow Jones series itself is not random noise, due to the upward trend, so we form the differences

```
index<-dow$Dow;
diff.dow<-diff(index);
plot.ts(diff.dow,col='blue',
ylab='first difference', main='Time Series
Plot of Differences')
abline(h=mean(diff.dow),col="red");
```

Time Series Plot of Differences



Dow Differences (cont.)

```
our.test<-runs(diff.dow);  
  
names(our.test);  
  
## [1] "T"          "Ta"         "Tb"         "R"          "Z"          "E(R)"
```

Dow Differences (cont.)

```
our.test$R;  
  
## [1] 26  
  
our.test$Z;  
  
## [1] 0.01144965  
  
our.test$"E(R)"  
  
## [1] 25.96  
  
our.test$"P value";  
  
## [1] 0.9908647
```

Dow Differences (cont.)

It appears to be a much more random series, varying around the mean difference 26. The runs test shows that there is absolutely no evidence of nonrandom differences.

Forecast of April 1992

Assuming the random walk model is adequate, the forecast of April 1992 made in March 1992 is the observed March value, 3247.42, plus the mean difference, 26, or 3273.42. A measure of the forecast accuracy is provided by the standard deviation, $s_D=84.65$, of the differences. Provided that our assumptions hold, we can be 95 % confident that our forecast is off by no more than $2s_D$, or about 170.

Exercise

Write a function in R that gives you a one-step-ahead forecast for a Random Walk Model.

One solution

```
forecast<-function(y){  
  
  diff<-diff(y)  
  y.diff.bar<-mean(diff);   #average difference;  
  last<-length(y); #last observation;  
  F.next<-y[last]+y.diff.bar;  
  new.y<-c(y,F.next);  
  
  list('Y(t)'=y, 'Y(t+1)'=new.y, 'F(t+1)'=F.next)  
  
}
```


Exercise

Using your one-step-ahead forecast function, write another function in R that computes forecast for times: $t+1, t+2, \dots, t+N$, where t represents the length of your original time series.

One solution

```
forecast.N<-function(y,N){  
  
original<-y;      #original time series;  
  
for (i in 1:N){  
new<-forecast(original)$'Y(t+1)' ;  
original<-new;  
      }  
  
list('Y(t) '=y, 'N' =N, 'Y(t+N)' =original)  
  
      }
```

Autoregression Models

A regression-based extrapolation method is to regress the current value of the time series on past (lagged) values.

This is called autoregression, where the "auto" means that the explanatory variables in the equation are lagged values of the response variable, so that we are regressing the response variable on lagged versions of itself. Some trial and error is generally required to see how many lags are useful in the regression equation. The following exercise illustrates the procedure.

Exercise: Hammers

A retailer has recorded its weekly sales of hammers (units purchased) for the past 42 weeks. (See the file `hammers.txt`) How useful is autoregression for modeling these data and how would it be used for forecasting?

Exercise: Hammers (cont.)

- Plot your time series
- Make a correlogram
- Fit an autoregressive model of order p (p suggested by your correlogram, pacf and/or aic)
- Determine if your model is adequate
- Fit another model if necessary

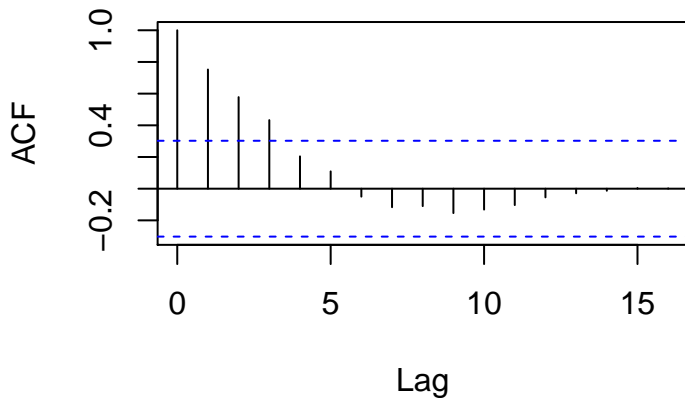
```
hammers<-read.table(file='hammers.txt',header=TRUE)

sales<-hammers$Sales

new.sales<-ts(sales)

acf(new.sales)
```

Series new.sales



```
y<-new.sales;  
  
m<-length(y)-3;  
  
y.lag1<-y[-1][1:m];  
  
y.lag2<-y[-c(m-1,m)][1:m];  
  
y.lag3<-y[-c(m-2,m-1,m)][1:m];  
  
response<-y[-c(1,2,3)];  
  
mod.ar3<-lm(response~y.lag1+y.lag2+y.lag3);  
  
summary(mod.ar3);
```



```
##
## Call:
## lm(formula = response ~ y.lag1 + y.lag2 + y.lag3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37.095  -9.759   2.334  10.755  33.934
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.1770     7.9834   3.028  0.0046 **
## y.lag1        0.4417     0.1839   2.402  0.0217 *
## y.lag2       -1.7073     0.7961  -2.144  0.0390 *
## y.lag3        1.8523     0.7585   2.442  0.0198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod<-ar(new.sales,method="ols",order.max=3);
mod;

##
## Call:
## ar(x = new.sales, order.max = 3, method = "ols")
##
## Coefficients:
##      1
## 0.7932
##
## Intercept: -0.3897 (2.355)
##
## Order selected 1  sigma^2 estimated as  227
```

Regression-Based Trend Models

Many time series follow a long-term trend except for random variation. This trend can be upward or downward. A straightforward way to model this trend is to estimate a regression equation for Y_t , using time t as the single explanatory variable. In this "section" we will discuss the two most frequently used trend models, linear trend and exponential trend.

A linear trend means that the time series variable changes by a constant amount each time period. The relevant equation is

$$Y_t = \alpha + \beta t + \epsilon_t \quad (11)$$

where, as in previous regression equations, α is the intercept, β is the slope, and ϵ_t is an error term.

Exercise: Reebok

The file `reebok.txt` includes quarterly sales data for Reebok from the first quarter 1986 through second quarter 1996. Sales increase from \$ 174.52 million in the first quarter to \$ 817.57 million in the final quarter. How well does a linear trend fit these data? Are the residuals from this fit random?

Exercise: Reebok (cont.)

- Plot your time series
- Fit a linear regression model and interpret your results
- Find the residuals
- Determine if the residuals are random

```
reebok<-read.table(file="reebok.txt",header=TRUE)

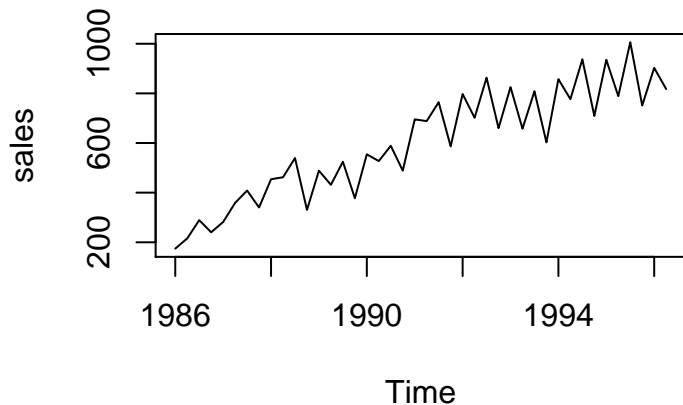
sales<-reebok$Sales

time<-reebok$Time

new.sales<-ts(sales,start=c(1986,1),end=c(1996,2),freq=4)

### Time Series Chart

plot(new.sales,ylab='sales')
```



Fitting Linear Model

```
mod<-lm(sales~time)
```

```
summary(mod)
```

```
##
```

```
## Call:
```

```
## lm(formula = sales ~ time)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

##	-171.218	-68.592	4.722	65.386	172.074
----	----------	---------	-------	--------	---------

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

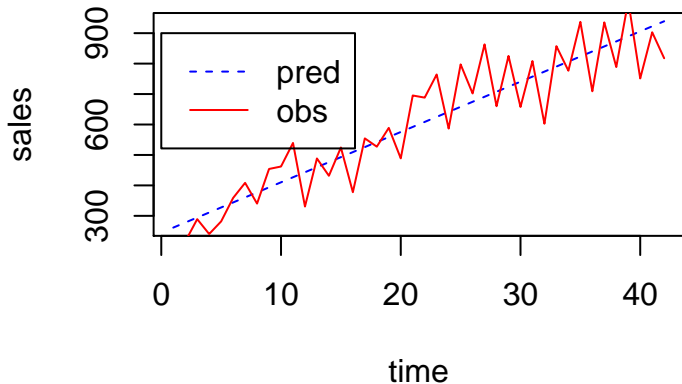
## (Intercept)	244.816	28.399	8.621	1.16e-10 ***
## time	16.530	1.151	14.366	< 2e-16 ***

```
##
```

Time Series Plot with Linear Trend Superimposed

```
plot(predict(mod), type='l', col='blue', xlab='time', ylab='sales')  
  
lines(time, sales, col='red');  
  
legend(0, 900, c('pred', 'obs'), col=c('blue', 'red'), lty=c(2, 1));
```

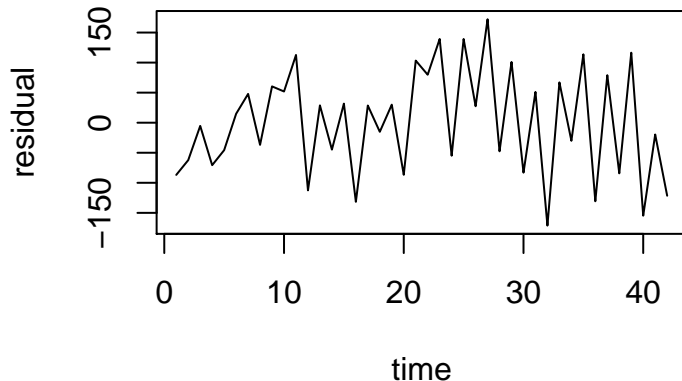
Time Series Plot with Linear Trend Superimposed



Residuals

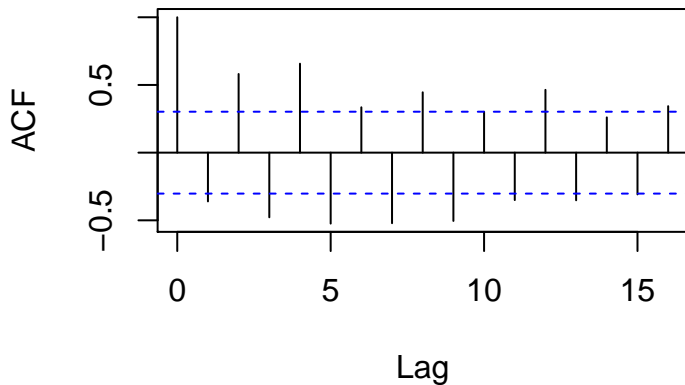
```
res<-mod$residuals;  
# residuals from our model;  
  
res<-ts(res);  
  
plot(res,xlab='time',ylab='residual',main='Time Series Plot of
```

Time Series Plot of Forecast Errors



```
acf(res);
```

Series res



Exponential Trend

In contrast to a linear trend, an exponential trend is appropriate when the time series changes by a constant percentage (as opposed to a constant dollar amount) each period. Then the appropriate regression equation is

$$Y_t = c \exp(bt)u_t \quad (12)$$

where c and b are constants, and u_t represents a multiplicative error term. By taking logarithms of both sides, and letting $a = \ln(c)$ and $\epsilon_t = \ln(u_t)$, we obtain a linear equation that can be estimated by the usual linear regression method. However, note that the response variable is now the logarithm of Y_t :

$$\ln(Y_t) = a + bt + \epsilon_t \quad (13)$$

Exercise: Intel

The file intel.txt contains quarterly sales data for the chip manufacturing firm Intel from the beginning of 1986 through the second quarter of 1996. Each sales value is expressed in millions of dollars. Check that an exponential trend fits these sales data fairly well. Then estimate the relationship and interpret it.

Exercise: Intel (cont)

- Plot your time series
- If your original time series shows an exponential trend, apply \ln to the series
- Use the transformed series to estimate the relationship
- Express the estimated relationship in the original scale
- Find an estimate of the standard deviation

Moving Averages

Perhaps the simplest and one of the most frequently used extrapolation methods is the method of moving averages. To implement the moving averages method, we first choose a span, the number of terms in each moving average. Let's say the data are monthly and we choose a span of 6 months. Then the forecast of next month's value is the average of the most recent 6 month's values. For example, we average January-June to forecast July, we average February-July to forecast August, and so on. This procedure is the reason for the term moving averages.

Exponential Smoothing

There are two possible criticisms of the moving averages method. First, it puts equal weight on each value in a typical moving average when making a forecast. Many people would argue that if next month's forecast is to be based on the previous 12 months' observations, then more weight ought to be placed on the more recent observations. The second criticism is that moving averages method requires a lot of data storage. This is particularly true for companies that routinely make forecasts of hundreds or even thousands of items. If 12-month moving averages are used for 1000 items, then 12000 values are needed for next month's forecasts. This may or may not be a concern considering today's relatively inexpensive computer storage capabilities.

Exponential Smoothing (cont.)

Exponential smoothing is a method that addresses both of these criticisms. It bases its forecasts on a weighted average of past observations, with more weight put on the more recent observations, and it requires very little data storage.

There are many versions of exponential smoothing. The simplest is called simple exponential smoothing. It is relevant when there is no pronounced trend or seasonality in the series. If there is a trend but no seasonality, then Holt's method is applicable. If, in addition, there is seasonality, then Winter's method can be used.

Simple Exponential Smoothing

We now examine simple exponential smoothing in some detail. We first introduce two new terms. Every exponential model has at least one smoothing constant, which is always between 0 and 1. Simple exponential smoothing has a single smoothing constant denoted by α . The second new term is L_t , called the level of the series at time t . This value is not observable but can only be estimated. Essentially, it is where we think the series would be at time t if there were no random noise.

Simple Exponential Smoothing (cont.)

Then the simple exponential smoothing method is defined by the following two equations, where F_{t+k} is the forecast of Y_{t+k} made at time t :

$$L_t = \alpha Y_t + (1 - \alpha)L_{t-1} \quad (14)$$

$$F_{t+k} = L_t \quad (15)$$

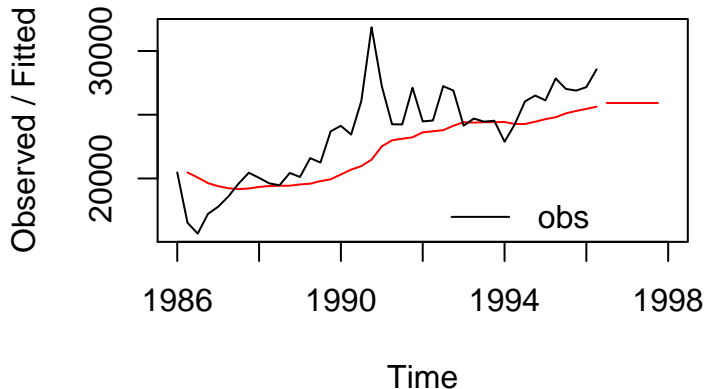
Example: Exxon

The file `exxon.txt` contains data on quarterly sales (in millions of dollars) for the period from 1986 through the second quarter of 1996. Does a simple exponential smoothing model track these data well? How do the forecasts depend on the smoothing constant α ?

Example: Exxon (R code)

```
exxon<-read.table(file="exxon.txt",  
header=TRUE);  
sales<-exxon$Sales;  
new.sales<-ts(sales,start=c(1986,1),  
end=c(1996,2),freq=4);  
mod1<-HoltWinters(new.sales,alpha=0.1,  
beta=FALSE,gamma=FALSE);  
plot(mod1,xlim=c(1986,1998));  
lines(predict(mod1,n.ahead=6),col="red");  
legend(1992,20000,c("obs","forecasts"),  
col=c("black","red"),lty=c(1,1),bty="n");
```

Holt-Winters filtering



Exxon (changing α)

The following lines of code produce a "movie" that illustrates the effect of changing α .

```
n<-100
for (i in 1:n){
  mod1<-HoltWinters(new.sales,
    alpha=(1/n)*i,beta=FALSE,gamma=FALSE)
  plot(mod1,xlim=c(1986,1997))
  Sys.sleep(0.3)
}
```

Choosing α .

What value of α should we use? There is no universally accepted answer to this question. Some practitioners recommend always using a value around 0.1 or 0.2. Others recommend experimenting with different values of α until a measure such as the Mean Square Error (MSE) is minimized. R can find this optimal value of α as follows:

```
mod2<-HoltWinters(new.sales,  
beta=FALSE,gamma=FALSE);
```

Holt's Model for Trend

The simple exponential smoothing model generally works well if there is no obvious trend in the series. But if there is a trend, then this method consistently lags behind it. Holt's method rectifies this by dealing with trend explicitly. In addition to the level of the series, L_t , Holt's method includes a trend term, T_t , and a corresponding smoothing constant β . The interpretation of L_t is exactly as before.

Holt's Model for Trend (cont.)

The interpretation of T_t is that it represents an estimate of the change in the series from one period to the next. The equations for Holt's model are as follows:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (16)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (17)$$

$$F_{t+k} = L_t + kT_t \quad (18)$$

Example: Dow Jones revisited

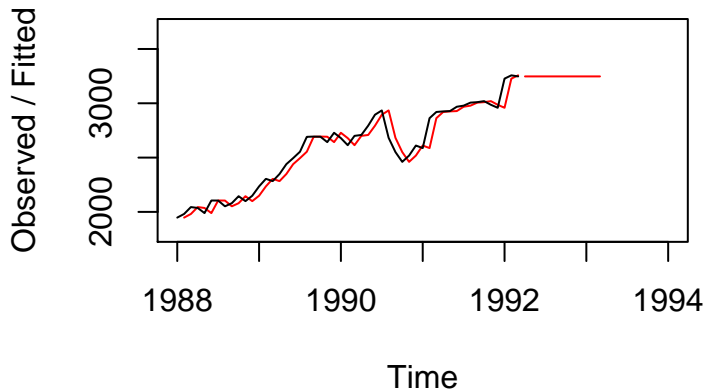
We return to the Dow Jones data (see file dow.txt). Again, these are average monthly closing prices from January 1988 through March 1992. Recall that there is a definite upward trend in this series. In this example we investigate whether simple exponential smoothing can capture the upward trend. Then we see whether Holt's exponential smoothing method can make an improvement.

Example: Dow Jones (R code)

```
dow<-read.table(file="dow.txt",header=TRUE);
index<-dow$Dow;
DJI<-ts(index,start=c(1988,1),end=c(1992,3),
freq=12);
mod1<-HoltWinters(DJI,beta=FALSE,gamma=FALSE);

### predictions
plot(mod1,xlim=c(1988,1994),
ylim=c(1800,3700));
preds<-predict(mod1,n.ahead=12);
lines(preds,col="red");
```

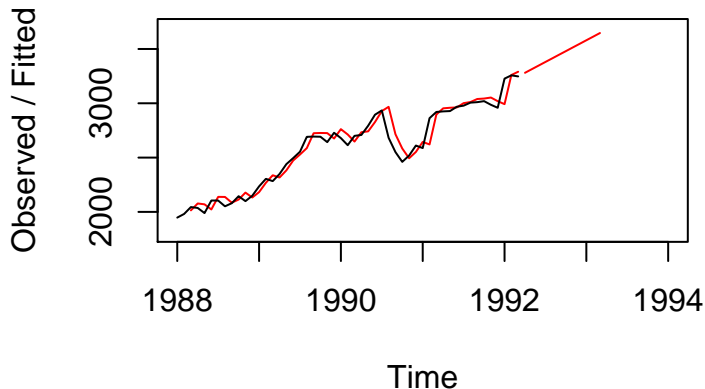

Holt-Winters filtering



R code (Holt's Model)

```
mod2<-HoltWinters(DJI,gamma=FALSE);  
#Fitting Holt's model;  
  
### predictions;  
  
plot(mod2,xlim=c(1988,1994),  
ylim=c(1800,3700));  
preds<-predict(mod2,n.ahead=12);  
lines(preds,col='red');
```

Holt-Winters filtering



So far we have said practically nothing about seasonality. Seasonality is defined as the consistent month-to-month (or quarter-to-quarter) differences that occur each year. For example, there is seasonality in beer sales - high in the summer months, lower in other months.

How do we know whether there is seasonality in a time series? The easiest way is to check whether a plot of the time series has a regular pattern of ups and /or downs in particular months or quarters.

Winter's Model (cont.)

Winter's exponential smoothing model is very similar to Holt's model- it again has a level and a trend terms and corresponding smoothing constants α and β - but it also has seasonal indexes and a corresponding smoothing constant γ . This new smoothing constant γ controls how quickly the method reacts to perceived changes in the pattern of seasonality.

Example: Coke

The data in the coke.txt file represent quarterly sales (in millions of dollars) for Coca Cola from quarter 1 of 1986 through quarter 2 of 1996. As we might expect, there has been an upward trend in sales during this period, and there is also a fairly regular seasonal pattern, as shown in the next slide. Sales in the warmer quarters, 2 and 3, are consistently higher than in colder quarters, 1 and 4. How well can Winter's method track this upward trend and seasonal pattern?

```
coke<-read.table(file="coke.txt",header=TRUE);
sales<-coke$Sales;
new.sales<-ts(sales,start=c(1986,1),end=c(1996,2),freq=4)
mod.coke<-HoltWinters(new.sales)
# Holt-Winter's model

### predictions

plot(mod.coke,ylim=c(1000,6000),
xlim=c(1986,1998))

preds<-predict(mod.coke,n.ahead=6)

lines(preds,col='red')
```

Holt-Winters filtering

