- The basic decomposition model for a time series X_t can be written as:

$$X_t = m_t + s_t + Y_t, \quad t = 1, 2, \ldots, T,$$

where $E[Y_t] = 0$, $s_{t+d} = s_t$, $\sum_{j=1}^{d} s_j = 0$ and the seasonal component s has period of length d.

- The first problem one faces is to estimate m_t and s_t at time point t.

- This is equivalent to smoothing or de-noising the data. That is getting rid of the the noise Y_t.

- A related problem is forecasting future values from the observed data.

- We first consider forecasting using the decomposition model.

- That is based on X_1, X_2, ..., X_T we try to predict the value of X_{T+k} by proposing $\widehat{X}_{T+k}$ for some positive integer k.

Naive Forecaster

- The most naive way to forecast the future values is to declare them to be equal to X_T, ie. the last observed value.

- The naive estimator does exactly that.

- That is for any $k \geq 1$ it takes:

$$\hat{X}_{T+k} = X_T.$$

- Notice that the predicted values of the naive estimator assumes the trend to be flat or constant.

- If that is not true the naive estimator would be an extremely bad predictor.

- In practice the naive forecaster is seldom used. Its main utility is in being a benchmark for other forecasting methods.

MOVING AVERAGE METHODS WITH p TERMS

- A moving average method is mainly a method for smoothing datasets.

- Several moving average methods are available.

- The one which is usually used in forecasting computes:

$$\widehat{X}_{pT} = \sum_{j=0}^{p-1} a_j X_{T-j} = a_0 X_T + a_1 X_{T-1} + \cdots + a_{p-1} X_{T-(p-1)}.$$

  Here a_i, $i = 0, 1, 2, \ldots, p-1$ are pre-specified constants such that $\sum_{j=0}^{p-1} a_j = 1$.

- If the trend is flat, for any $k \geq 1$ the k-step ahead forecast value of X_{T+k} is given by

$$\widehat{X}_{T+k} = \widehat{X}_{pT}.$$

- The choices of p and a_i, $i = 0, 1, 2, \ldots, p$ depend on the nature of the problem.

- With a small p one may get "jagged" curves. On the other hand a large p may over-smooth the data.

Linear Moving Average method

- Most simple and probably most widely used linear moving average method takes $a_i = 1/p$ for all $i = 0, 2, \ldots, p-1$.

- That is $\widehat{X}_{pT}$ is the mean of last p observations.

$$\widehat{X}_{pT} = \left(X_T + X_{T-1} + \cdots + X_{T-(p-1)}\right)/p.$$

- As before, if the trend is flat, for any $k \geq 1$, we can forecast

$$\widehat{X}_{T+k} = \widehat{X}_{pT}.$$

- Clearly, the choice of p is crucial. A small p will "under-smooth" and a large p will over-smooth the series.

- Two popular choices of p are 3 and 12.

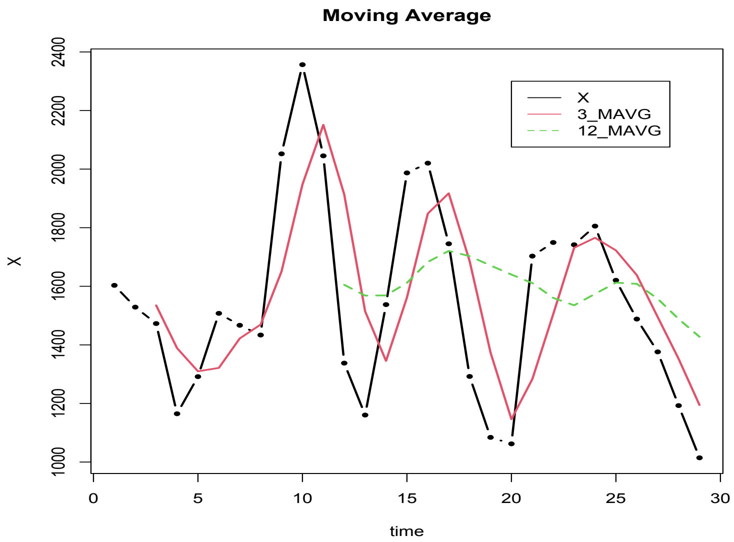
Example

Comparison of a Three-Period and 12-Period Moving Average on an Annual Housing Starts Series

Datum Numbers	Housing Starts (1963–1991)	3_MAVG	12_MAVG
1	1603.2		
2	1528.8		
3	1472.8	1534.9	
4	1164.9	1388.8	
5	1291.6	1309.8	
6	1507.6	1321.4	
7	1466.8	1422.0	
8	1433.6	1469.3	
9	2052.2	1650.9	
10	2356.6	1947.5	
11	2045.3	2151.4	
12	1337.7	1913.2	1605.1
13	1160.4	1514.5	1568.2
14	1537.5	1345.2	1568.9
15	1987.1	1561.7	1611.8
16	2020.3	1848.3	1683.1
17	1745.1	1917.5	1720.9
18	1292.2	1685.9	1702.9
19	1084.2	1373.8	1671.0
20	1062.2	1146.2	1640.1
21	1703.0	1283.1	1611.0
22	1749.5	1504.9	1560.4
23	1741.8	1731.4	1535.1
24	1805.4	1765.6	1574.1
25	1620.5	1722.6	1612.4
26	1488.1	1638.0	1608.3
27	1376.1	1494.9	1557.4
28	1192.7	1352.3	1488.4
29	1014.5	1194.4	1427.5
30		1194.4	1427.5
31		1194.4	1427.5
32		1194.4	1427.5
33		1194.4	1427.5

(*Source*: Dielman, 1996, Ex 3.4 HSTARTS3.DAT)

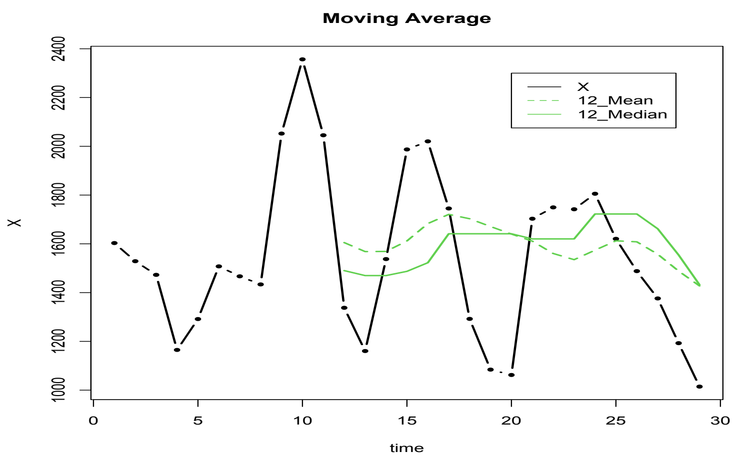
EXAMPLE (CONTD.



- The 12 year moving average gives a smoother fit than the 3 year moving average.
- The typical forecast of the 12 year moving average seems to be more closer to the centre of the data than the 3 year one.
- The 3 year moving average shows more cyclicity than the 12 year one. The former follows the peaks and troughs of the historical data, but the cyclical pattern will lag behind the original one.

MOVING MEDIAN METHOD

- Moving median method uses median of the last p observations instead of their mean.
- There are advantages and disadvantages in using the median instead of the mean.
- The advantage is that the median is more robust to outliers and can handle them better.
- However, medians don't smooth that well. After the smoothing the curve may still look quite "jagged".



Exponential Smoothing Method

- It is intuitive that the observations in recent past would be more relevant in forecasting than the observations of distant past.
- Thus one would like to put more weight on the recent observations.
- This sort of smoothing methods are called exponential smoothing.
- The exponential smoothing uses the whole history and smoothes the whole dataset.
- The steps are as follows:
  1. Choose a $0 \leq \alpha \leq 1$.
  2. $\widehat{X}_{h1} = X_1$,
  3. $\widehat{X}_{ht} = \alpha X_t + (1-\alpha)\widehat{X}_{h(t-1)}$ for $t = 2, 3, \ldots, T$.
- Note that in exponential smoothing $\widehat{X}_{ht}$ is a convex combination of the observed value at time t and the smoothed value at time $t-1$.
- It is not difficult to see that $\widehat{X}_{h2} = \alpha X_2 + (1-\alpha)X_1$, $\widehat{X}_{h3} = \alpha X_3 + \alpha(1-\alpha)X_2 + (1-\alpha)^2 X_1$ etc.

Exponential Smoothing Method (contd.)

- In general for any $t \geq 2$,

$$\widehat{X}_{ht} = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1.$$

- It can be also be seen that exponential smoothing is a moving average method.

- For forecasting the value at $T+1$ when the trend is flat we use as before use:

$$\widehat{X}_{T+1} = \widehat{X}_{hT}.$$

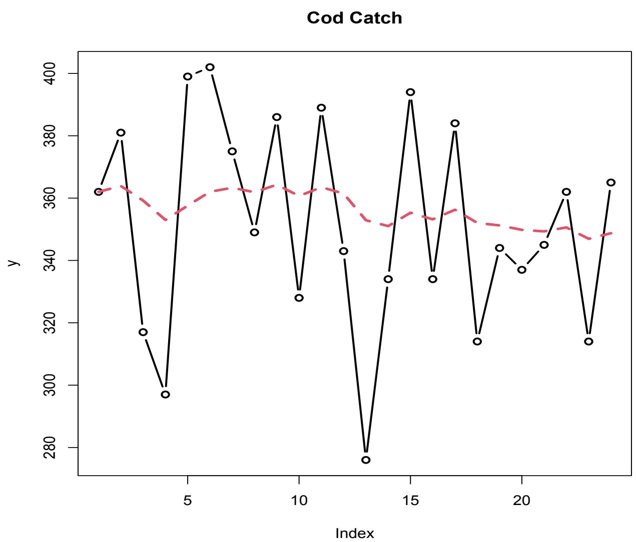
- Notice that exponential smoothing uses weights monotonically decreasing over lag. Consider the table below:

Weight assigned to:	$\alpha = .1$	$\alpha = .3$	$\alpha = .5$	$\alpha = .9$
X_T	0.1	0.3	0.5	0.9
X_{T-1}	0.09	0.21	0.25	0.09
X_{T-2}	0.081	0.147	0.125	0.009
X_{T-3}	0.0729	0.1029	0.0625	0.0009

EXAMPLE

	Data	Fitted Values	Error	Squared Error
1	362.00	362.00	0.00	0.00
2	381.00	363.90	17.10	292.41
3	317.00	359.21	-42.21	1781.68
4	297.00	352.99	-55.99	3134.77
5	399.00	357.59	41.41	1714.78
6	402.00	362.03	39.97	1597.51
7	375.00	363.33	11.67	136.24
8	349.00	361.90	-12.90	166.29
9	386.00	364.31	21.69	470.64
10	328.00	360.68	-32.68	1067.66
11	389.00	363.51	25.49	649.86
12	343.00	361.46	-18.46	340.65
13	276.00	352.91	-76.91	5915.32
14	334.00	351.02	-17.02	289.68
15	394.00	355.32	38.68	1496.29
16	334.00	353.19	-19.19	368.11
17	384.00	356.27	27.73	769.09
18	314.00	352.04	-38.04	1447.11
19	344.00	351.24	-7.24	52.37
20	337.00	349.81	-12.81	164.18
21	345.00	349.33	-4.33	18.76
22	362.00	350.60	11.40	129.99
23	314.00	346.94	-32.94	1084.96
24	365.00	348.74	16.26	264.23

	Year 1	Year 2
January	362.00	276.00
February	381.00	334.00
March	317.00	394.00
April	297.00	334.00
May	399.00	384.00
June	402.00	314.00
July	375.00	344.00
August	349.00	337.00
September	386.00	345.00
October	328.00	362.00
November	389.00	314.00
December	343.00	365.00



SIMPLE EXPONENTIAL SMOOTHING

1. Suppose that the time series $y_1, y_2, \ldots, y_n$ has a level (or mean) that may be slowly changing over time but has no trend or seasonal pattern. Then the estimate ℓ_T for the level (or mean) of the time series in time period T is given by the **smoothing equation**

$$\hat{X}_{hT} = \alpha X_T + (1-\alpha)\hat{X}_{h(T-1)}$$

where α is a **smoothing constant** between 0 and 1, and ℓ_{T-1} is the estimate of the level (or mean) of the time series in time period $T-1$.

2. A **point forecast made in time period T for $y_{T+\tau}$** is

$$\hat{X}_{T+\tau}(T) = \hat{X}_{hT} \quad (\tau = 1, 2, 3, \ldots)$$

3. If $\tau = 1$, then a 95% prediction interval computed in time period T for y_{T+1} is

$$[\hat{X}_{hT} \pm z_{[.025]}s]$$

If $\tau = 2$, then a 95% prediction interval computed in time period T for y_{T+2} is

$$\left[\hat{X}_{hT} \pm z_{[.025]}s\sqrt{1+\alpha^2}\right]$$

In general for any τ, **a 95% prediction interval computed in time period T for $y_{T+\tau}$** is

$$\left[\hat{X}_{hT} \pm z_{[.025]}s\sqrt{1+(\tau-1)\alpha^2}\right]$$

where the standard error s at time T is

$$s = \sqrt{\frac{\text{SSE}}{T-1}} = \sqrt{\frac{\sum_{t=1}^{T}[X_t - \hat{X}_{ht}(t-1)]^2}{T-1}} = \sqrt{\frac{\sum_{t=1}^{T}[y_t - \hat{X}_{h(t-1)}]^2}{T-1}}$$

Note: There is not general agreement on dividing the SSE by (T – number of smoothing constants). However, we use this divisor because it agrees with the computation of s in the equivalent Box–Jenkins models in Chapters 9 to 12.

$\hat{X}_{25} = 348.6385.$

$\hat{X}_{26} = 348.6385.$

$\vdots$

- Notice that as τ increases the width of the prediction interval increases as well.
- This is natural. One should be more uncertain in predicting ~~the futur~~ more distant future.

Smoothing profiles for Trend and Seasonality

- Exponential smoothing techniques are useful in forecasting when the data shows trend and seasoanlity.
- Pegels classified 12 forecasting profiles for exponential smoothing.

Pegels Classification of Exponential Smoothing Techniques Extended to Include the Damped Trend Technique