Chapter 6

Dr. Joel Fingerman, Justin Shea

July 6, 2016

The preceding three chapters of this book dealt with regression methods applied to cross sectional data. When we have cross sectional data our approach to forecasting is to use the explanatory variables to forecast the value of the dependent variable. This approach to forecasting is often termed *interpolation.*

The other broad method of quantitative business forecasting is to use the history of the data we wish to forecast. We examine the historical pattern of the data we wish to forecast and try to find a discernible pattern. Using this pattern, we forecast forward the future pattern of the data. This is method of forecasting is often termed *extrapolation.* The use of the historical pattern of the data to be forecast is known as *time series methods.*

Stage 1 Collection and Analysis of Times Series Data

"The future lies ahead."

— Mort Sahl

A time series of data is data collected sequentially over equal periods of time. There are many forms of time series: time series can be collected as daily data (such as daily stock market data), weekly (such as weekly receipts data), monthly (such as monthly sales data), quarterly (such as quarterly revenue data), or yearly (such as annual profits data).

Remember though, the time period of data collection must be consistent. If our time series data are both in monthly and quarterly form then forecasting can be exceedingly difficult. When you collect time series data be certain that the time period of collection will remain consistent.

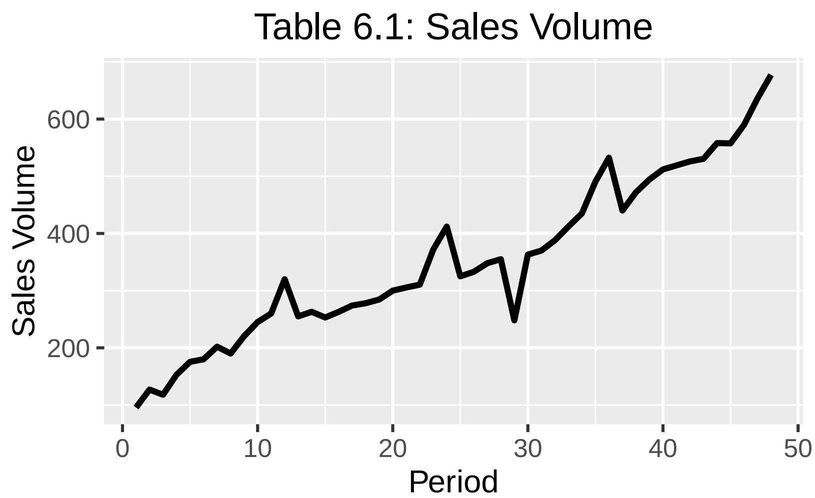
To create a time series of data from our example of Sales Volume data we now collect a series of Sales Volume data from *just one location over consecutive monthly periods*. We collected a data set of four years, 48 consecutive months (48 periods) of Sales Volume.

Table 1: Four Years of Time Series Data

Period	Year 1	Period	Year 2	Period	Year 3	Period	Year 4
1	95.5	13	255.0	25	325.0	37	440.0
2	127.0	14	263.0	26	333.0	38	472.0
3	118.0	15	253.0	27	348.0	39	494.5
4	153.0	16	263.0	28	355.0	40	512.0
5	175.5	17	274.0	29	248.0	41	519.0
6	180.0	18	278.0	30	363.0	42	526.0
7	202.0	19	284.5	31	370.0	43	530.5
8	190.0	20	300.0	32	388.0	44	558.0
9	220.5	21	305.5	33	412.0	45	557.5
10	245.0	22	310.5	34	435.0	46	590.0
11	260.0	23	372.0	35	490.5	47	636.4
12	320.0	24	412.0	36	532.5	48	677.4

Graphically displaying Time Series Data: Time Series Plot

The corresponding *time series plot* or *time series graph* of the Sales Volume is in Figure 6-1 below.



Tips on Graphing Times Series Data: The Practice

Good graphical presentations of time series data and forecasts can be very useful and very powerful in the presentation of ideas. A good layout for a time series graph is always helpful in the data analysis, model fitting, and forecasting.

Notice in our graph, that we allowed for additional space to the right of the last observation of our data series. We have 48 observations but allowed for 54 observations. Notice too that because the time series graph reveals an upward direction of the data series we also allowed for additional space above the last observation. The value of the last observation is 677.4 but allowed for values up to 800. In other words we have provided for plenty of "white space" for our graph.

Along the horizontal or time period axis we created horizontal reference lines at every six periods. It would be too dense to create horizontal reference lines at every of 54 periods. And it would be too sparse to form lines at every 12th period. These are monthly data so every 6 periods denotes a half year. Period 6 is a June, Period 12 is a December, Period 18 is June again, and so on.

Along the vertical or Sales Volume axis we set vertical reference lines at every 100 units. Units of 100 allow for sufficient spacing so the graph does not become crowded, and yet detailed enough so that we can estimate values with reasonable accuracy.

In summary, when producing time series graph, try to make the graphs uncluttered without sacrificing accuracy or information.

Time Series Data: Time Series Notation

When discussing observations in our data set we think of the data in the form

Past or historical observations
..., Y_{T-1}, Y_{T-2}, Y_{T-3}

Most recent observation
Y_T

Future observation
Y_{T+1}, Y_{T+2}, Y_{T+3}, ...

$$Y_1, Y_2, Y_3, ..., Y_T$$

Y_1 is the first observation, Y_2 is the second observation, and so on. is the oldest observation down through Y_T which is the most recent observation. In our example data series:

$$Y_1 = 95.5$$

$$Y_2 = 127.0$$

$$Y_3 = 118.0$$

$$\vdots$$

$$Y_{48} = 677.4$$

The uppercase letter T of Y_T will always mean the most recent observation in our time series. Y_{T+1} means then an observation one period into the future. Thus, Y_{T+1}, Y_{T+2}, Y_{T+3}, denote future observations one, two, and three periods beyond the present time Y_T. Similarly, Y_{T-1} denotes an observation one period in the past, so that Y_{T-1}, Y_{T-2}, Y_{T-3}, denote observations one, two, and three periods in the past.

We can generalize this notation to:

$Y_{T+\ell}$: A future observation of ℓ periods ahead.

$Y_{T-\ell}$: A historical observation of ℓ periods back.

As we distinguish between Y_i and $\hat{Y}_i$ in regression models, we shall destinguish between $Y_{T+\ell}$ and $\hat{Y_T}\ell$ in time series models.

$Y_{T+\ell}$ is the future observation of Y, ℓ period ahead, and unkown at time T.

$\hat{Y_T}\ell$ is the forecast of the future observation, ℓ periods ahead, made at time T.

Stage 1 Continued: The Theory of Collecting & Analyzing Data

We devote most of this chapter to Stage 1 of the Forecasting Process. With time series we recommend considerable investigation and analysis, to really understand the time series data, to really understand what the data are "telling us" before attempting to develop forecasting models. Consequently we shall begin this time series analysis with some easy, straightforward methods of analysis.

Smoothing Out Time Series Data

Almost all time series data are rough, having a stochastic element to the series. We shall discuss methods of "smoothing out" the data so that we may observe an underlying structure or pattern. The first, and simplest, method of smoothing time series data is through averaging.

Smoothing Out the Data by Averaging

If we believe there exists some underlying trend or pattern in the time series data, then by "smoothing the data" we smooth out (or average out) the random variations to reveal the underlying pattern in the series.

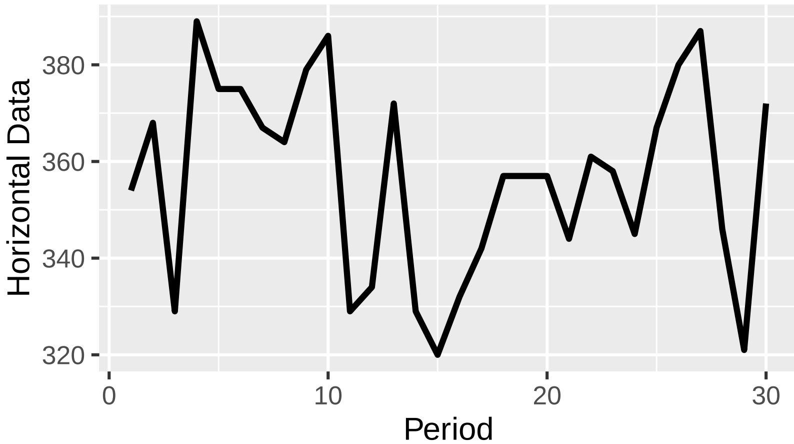
The Practice: Simple Average Smoothing

We begin with a time series of 30 observations, $T = 30$, which has a constant mean, or a slowly changing mean, over time, we call this "horizontal data". As in Table 6.2 and Figure 6.3.

Table 2: Horizontal Data: Table 6.2

Period	Actual	Period	Actual
1	354	16	332
2	368	17	342
3	329	18	357
4	389	19	357
5	375	20	357
6	375	21	344
7	367	22	361
8	364	23	358
9	379	24	345
10	386	25	367
11	329	26	380
12	334	27	387
13	372	28	346
14	329	29	321
15	320	30	372

Figure 6.3: Horizonal Data



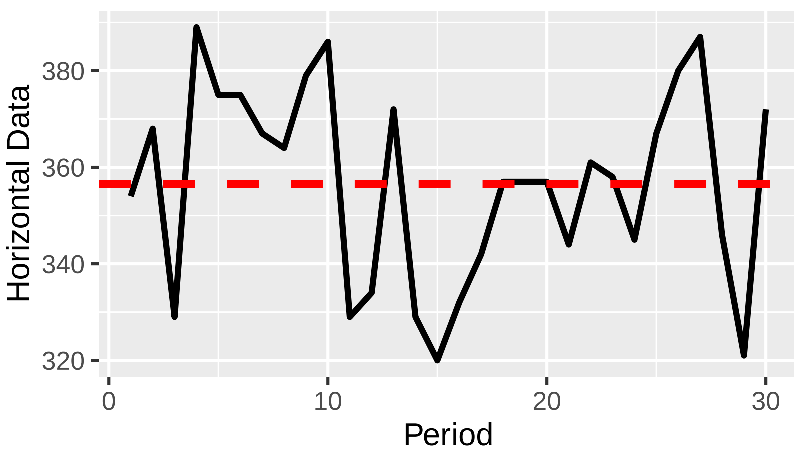
Becuase the Y_t are varying around a constant mean, a first smoothing of Y_t is just its mean, $\bar{Y}$[1]

In this example, the sample mean of $\bar{Y} = 356.5$.

[1] Smoothing using the sample mean

$$\bar{Y} = \sum_{t=1}^{T} Y_t$$

Horizonal Data with Plotted Mean



Moving Average Smoothings

The simple average discussed above is a smoothing method over all observations. A *moving average* is the technique of creating successive new averages by dropping the "oldest" observation and adding the most recent observation to calculate the new average.

3-period Centered Moving Average Smoothing for period t

To begin with an example, examine the notation in the margin[2]. Next, we shall use the data from Table 6.2 to illustrate this method. For example, a 3-period Centered Moving Average (at period 7) is:

[2] 3-period Centered Moving Average:

$$CMA_t(3) = \frac{Y_{T+1} + Y_T + Y_{T-1}}{3}$$

$CMA_7(3) = \frac{Y_8+Y_7+Y_6}{3}$
$CMA_7(3) = \frac{364+367+375}{3} = 368.67$

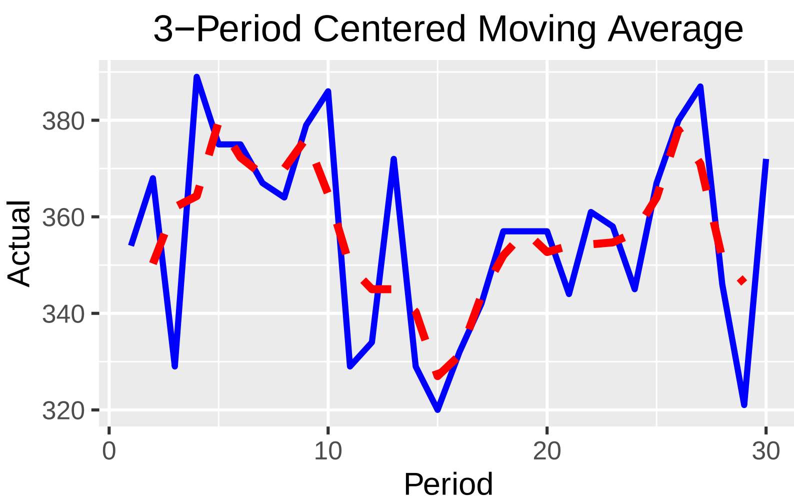
Table 3: Horizontal Data Table 6.2 subset: Centered Smoothing Average (3)

Period	Actual	Smoothing
6	375	...
7	367	368.67
8	364	...

This smoothing method is called a 3-period centered moving average because the computed value is placed at the center, or middle, of the 3 periods being used in the calculations.

Table 4: Table 6.3: Centered Smoothing Average (3)

Period	Actual	CMA(3)	Period	Actual	CMA(3)
1	354	NA	16	332	331.3
2	368	350.3	17	342	343.7
3	329	362.0	18	357	352.0
4	389	364.3	19	357	357.0
5	375	379.7	20	357	352.7
6	375	372.3	21	344	354.0
7	367	368.7	22	361	354.3
8	364	370.0	23	358	354.7
9	379	376.3	24	345	356.7
10	386	364.7	25	367	364.0
11	329	349.7	26	380	378.0
12	334	345.0	27	387	371.0
13	372	345.0	28	346	351.3
14	329	340.3	29	321	346.3
15	320	327.0	30	372	NA



5-period Centered Moving Average Smoothing for period t

A 5 period moving average smoothing is "smoother" than a 3 period because it uses a larger set of observations.[3]. As an example we smooth period 7 of the data in Table 6.3.

$CMA_7(5) = \frac{Y_9+Y_8+Y_7+Y_6+Y_5}{5}$
$CMA_7(5) = \frac{379+364+367+375+375}{5} = 372$

[3] 5-period Centered Moving Average:

$$CMA_t(5) = \frac{Y_{T+2} + Y_{T+1} + Y_T + Y_{T-1} + Y_{T-2}}{5}$$

Table 5: Horizontal Data Table 6.3 subset: Centered Smoothing Average (5)

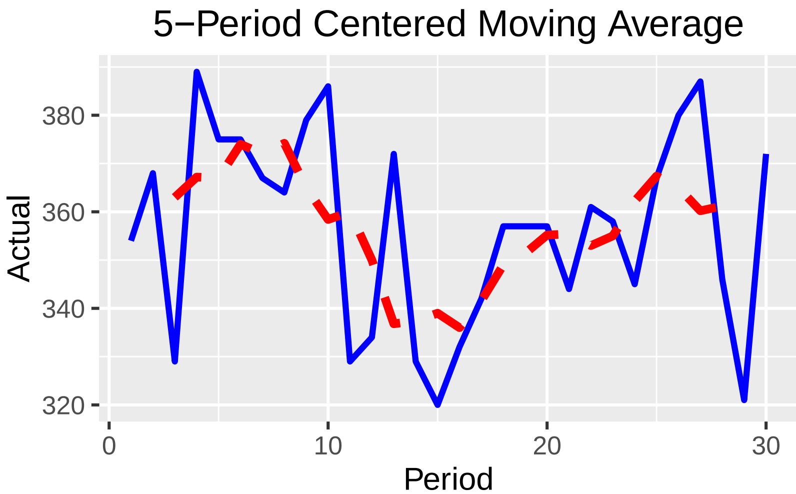
Period	Actual	Smoothing
4	389	...
5	375	...
6	375	...
7	367	372
8	364	...
9	379	...
10	386	...

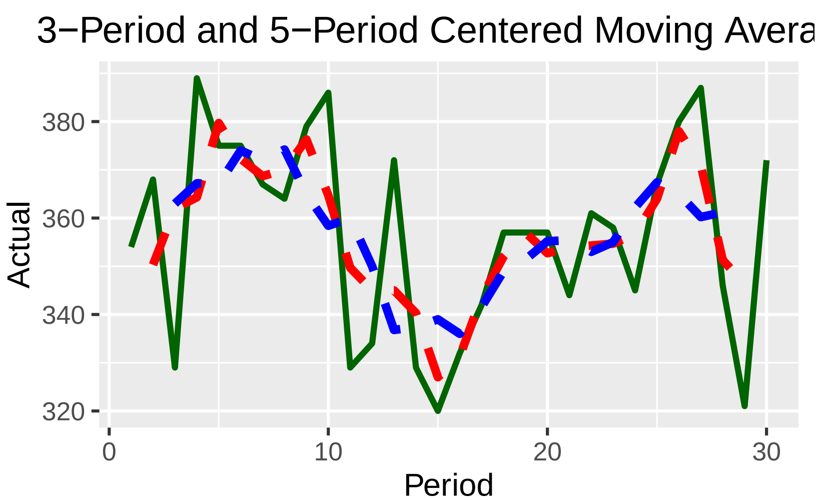
We list below both the 3-period and 5-period moving average smoothing and a graph of the actual values with the two smoothings superimposed.

Table 6: Table 6.4: Centered Smoothing Average (3) & (5)

Period	Actual	CMA(3)	CMA(5)
1	354	NA	NA
2	368	350.3	NA

Period	Actual	CMA(3)	CMA(5)
3	329	362.0	363.0
4	389	364.3	367.2
5	375	379.7	367.0
6	375	372.3	374.0
7	367	368.7	372.0
8	364	370.0	374.2
9	379	376.3	365.0
10	386	364.7	358.4
11	329	349.7	360.0
12	334	345.0	350.0
13	372	345.0	336.8
14	329	340.3	337.4
15	320	327.0	339.0
16	332	331.3	336.0
17	342	343.7	341.6
18	357	352.0	349.0
19	357	357.0	351.4
20	357	352.7	355.2
21	344	354.0	355.4
22	361	354.3	353.0
23	358	354.7	355.0
24	345	356.7	362.2
25	367	364.0	367.4
26	380	378.0	365.0
27	387	371.0	360.2
28	346	351.3	361.2
29	321	346.3	NA
30	372	NA	NA





4-Period Centered Moving Average Smoothing

A 4-period Centered Moving Average is possible, but the issue of the placing of the results remains.[4] Technically, if (say) the first 4 periods are used,

$CMA_t(4) = \frac{Y_4+Y_3+Y_2+Y_1}{4}$

then the placement of $CMA_t(4)$ is between periods 2 and 3 at "period 2.5."

$CMA_t(4) = \frac{389+329+368+354}{4} = 360$

The next 4-period centered smoothing is placed at "period 3.5"

$CMA_t(4) = \frac{Y_5+Y_4+Y_3+Y_2}{4}$

$CMA_t(4) = \frac{375+389+329+368}{4} = 365.25$

Then the "adjusted centered smoothing" for period 3 is the average of period 2.5 and period 3.5 smoothing.

$CMA_t(4) = \frac{360+365.25}{2} = 362.63$

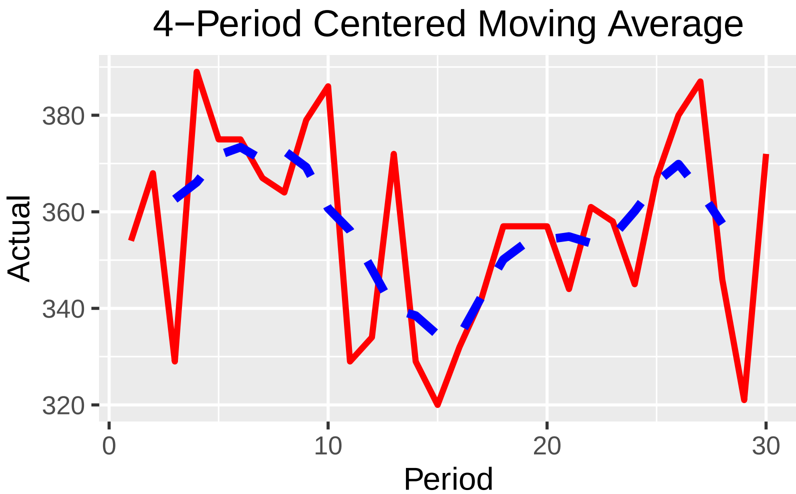
[4] 4-period Centered Moving Average:

$$CMA_t(4) = \frac{Y_{T+2} + 2Y_{T+1} + 2Y_T + 2Y_{T-1} + Y_{T-2}}{8}$$

Algebraically, it can be shown that the adjusted moving average of 4 periods, starting with period $t = 3$ is shown on the margin.

Table 7: Table 6.5: Centered Smoothing Average (4)

Period	Actual	CMA(4)
1	354	NA
2	368	NA
3	329	362.62
4	389	366.12
5	375	371.75
6	375	373.38
7	367	370.75
8	364	372.62
9	379	369.25
10	386	360.75
11	329	356.12
12	334	348.12
13	372	339.88
14	329	338.50
15	320	334.50
16	332	334.25
17	342	342.38
18	357	350.12
19	357	353.50
20	357	354.25
21	344	354.88
22	361	353.50
23	358	354.88
24	345	360.12
25	367	366.12
26	380	369.88
27	387	364.25
28	346	357.50
29	321	NA
30	372	NA



The 4-period Centered Moving Average Smoothing is especially suited for quarterly time series. However, the Centered Moving Average Smoothing formula may be generalized to any even number of periods.

References