

7. How to use Excel Forecast Sheet (ETS method)

A few years ago, Excel got a new routine that can be found under the Data tab and it is called Forecast Sheet. Fig. 7.1 shows the icon from the Excel ribbon bar.

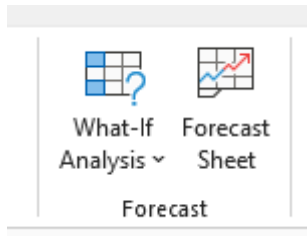


Fig 7.1

This routine is a complete forecasting suite based on a method called ETS, which stands for Error, Trend, Seasonality. The ETS method is a member of a much larger group of models, called state-space models. Fundamentally, it takes three components, which are the errors, a trend and seasonality and puts them in a particular relationship (additive or multiplicative). A specific model that has been implemented in Excel is the so-called AAA model. That's all you need to know.

The letter S in ETS implies that this is a particularly good method to tackle seasonal data. The stock values are not seasonal, but we'll see how to use this method for forecasting stocks and shares. Before we tackle the shares, let's use a simple non-stock exchange example just to explain the mechanics of the method.

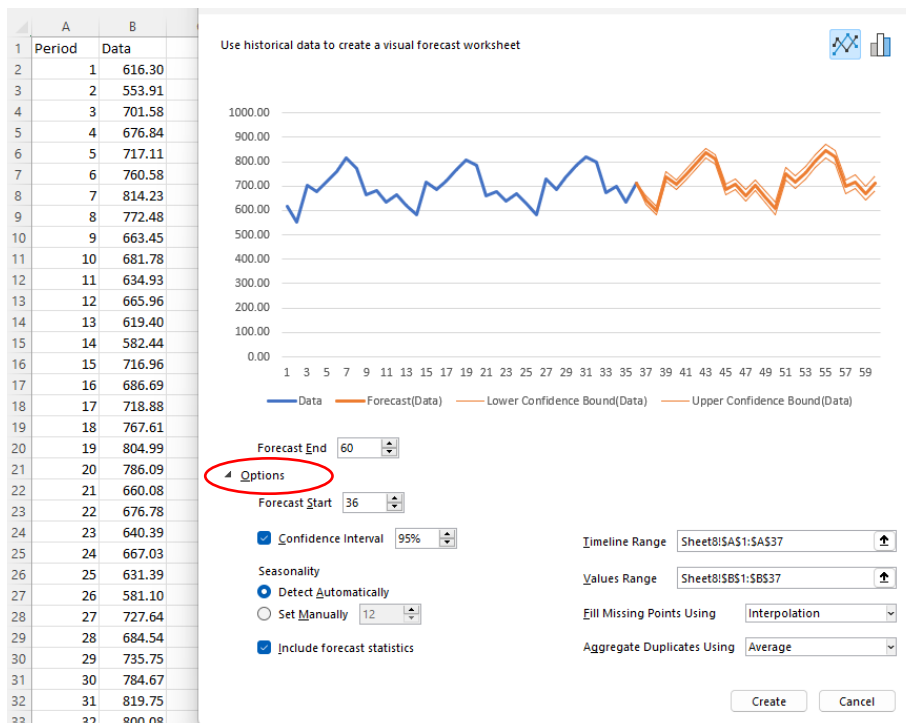


Fig. 7.2

In Fig. 7.2 we used a time series with only 36 observations (only the first 32 are shown). When you highlight the data in columns A and B, click on Forecast Sheet, as shown in Fig. 7.1.

What you will instantly see, as shown in Fig. 7.2, is a large dialogue box with the graph of the time series and instant forecasts, together with the prediction interval given. You should click on the Options button (circled in red in Fig. 7.1) and this will open the rest of the box, as shown here.

You can see that Excel automatically proposed forecasts until period 60, as well as where the forecasts should start from, though you can change this manually. Equally, the seasonality was in this case picked automatically, and Excel detected that it was a 12-period seasonality.

We also selected the button to Include forecast statistics. There are a few other options on the right-hand side of the box, but we can safely ignore them. The moment you click on the Create button, Excel opens a new sheet and gives you the full forecast. Fig. 7.3 shows the final result.

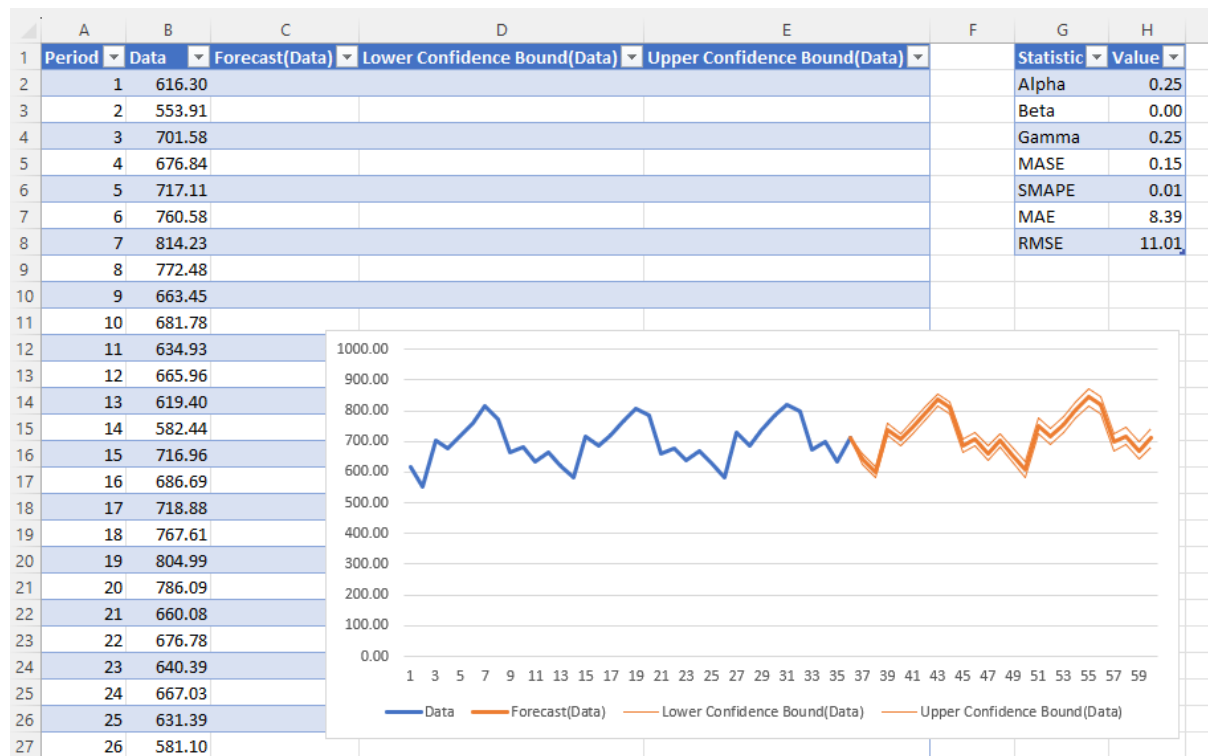


Fig. 7.3

Not shown in Fig. 7.3, but the rows go until 60 and show forecasts as well as the Lower and Upper Confidence interval. This is all depicted in the graph which is also automatically produced. You also get some statistics in the range G1:H8. The first three rows are the parameters of the model (which we will ignore) and the last four are different error metrics. We already met MAE (Mean Absolute Error) and RMSE (Root Mean Square Error), though MASE and SMAPE are new.

MASE stands for Mean Absolute Scaled Error and SMAPE stands for Symmetric Mean Absolute Percentage Error.

SMAPE is a relative error, so what you get are the numbers between 0 and 1. When you see 0, this implies a perfect match between the forecasted and actual values, while 1 means that there is no similarity between the two.

MASE, on the other hand, is a bit more complicated. It starts with the premise that today's value is the best predictor of tomorrow's value. This is called a naïve forecast. Then it measures how far is the forecast produced by ETS vs. the naïve forecast. A naïve forecast MASE metric has a value of 1.

Anything above 1 is NOT as good as a naïve forecast. Anything below 1 is better than the naïve forecast. That's all you need to know to interpret these two error metrics.

In our case SMAPE=0.01 and MASE=0.15. This means that according to both error metrics, we have a really, really good forecast.

If you select a cell C38 and drag it upwards, you effectively copy the forecast formula all the way to the beginning of the series and you can see how well it fits the data. Fig. 7.4 shows some of the functions and the results, as well as the graph.

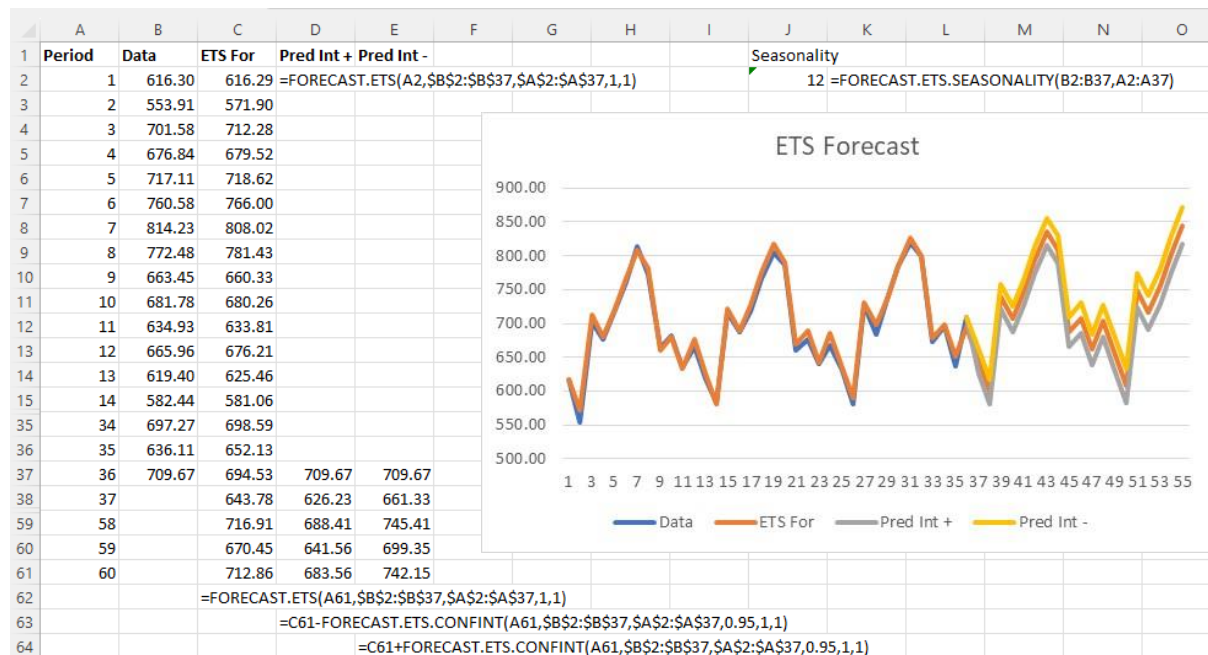


Fig. 7.4

From the graph in Fig. 7.4 you can hardly see the difference between the fitted values (in orange) and actual values (in blue). No wonder both MASE and SMAPE values were so good.

OK, now you know how to apply the Excel Forecast Sheet algorithm, let's see how to apply it to stocks and shares data.

Before we do, take a look at cell J2 in Fig. 7.4. It gives you the value of 12, which is the seasonality of this time series. As you can see, we got it using a dedicated Excel function:

`=FORECAST.ETS.SEASONALITY(B2:B37,A2:A37)`. This means that instead of selecting the seasonality to be calculated automatically, we could have used this function and then manually inserted the seasonality factor in `=FORECAST.ETS(A2,B2:B37,A2:A37,1,1)`, which is cell C1 and downwards. However, in this case, the function would look as follows: `=FORECAST.ETS(A2,B2:B37,A2:A37,12,1)`. Just park this piece of info for now.

As we said, there is no seasonality in stock data. There could be some cyclical variations, and because mathematically they are similar to seasonal variations, we will try to use this method to forecast some of our shares.

I picked HSBC (HSBA) closing daily values from 4 Jan 2010 until 29 Dec 2023. The graph is shown in Fig. 7.5. As we can see it seems to be showing some cyclical variations, but we'll come back to this point later. As a first step, let's just highlight the stock values in the spreadsheet and click on Forecasts Sheet as before.

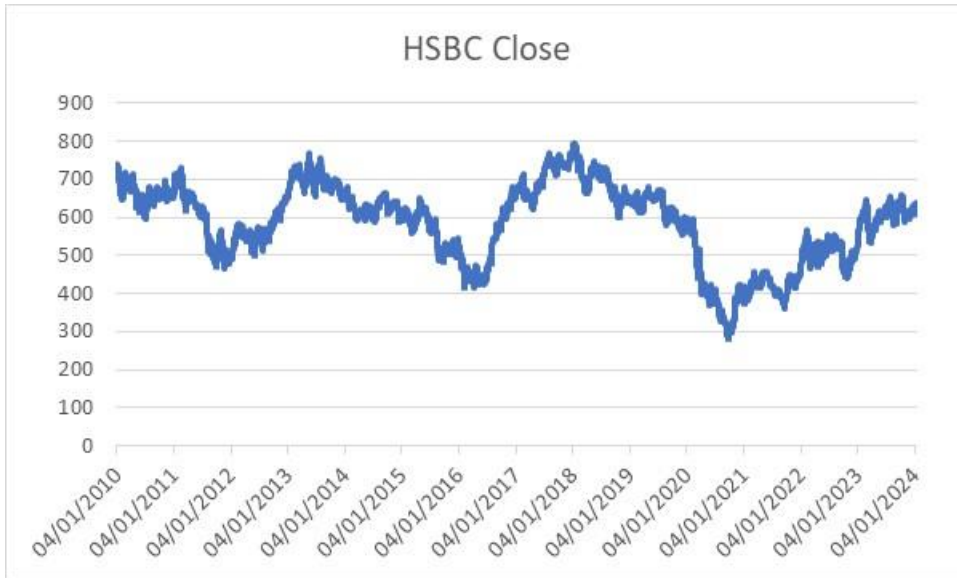


Fig. 7.5

What we get instantly is shown in Fig. 7.6.

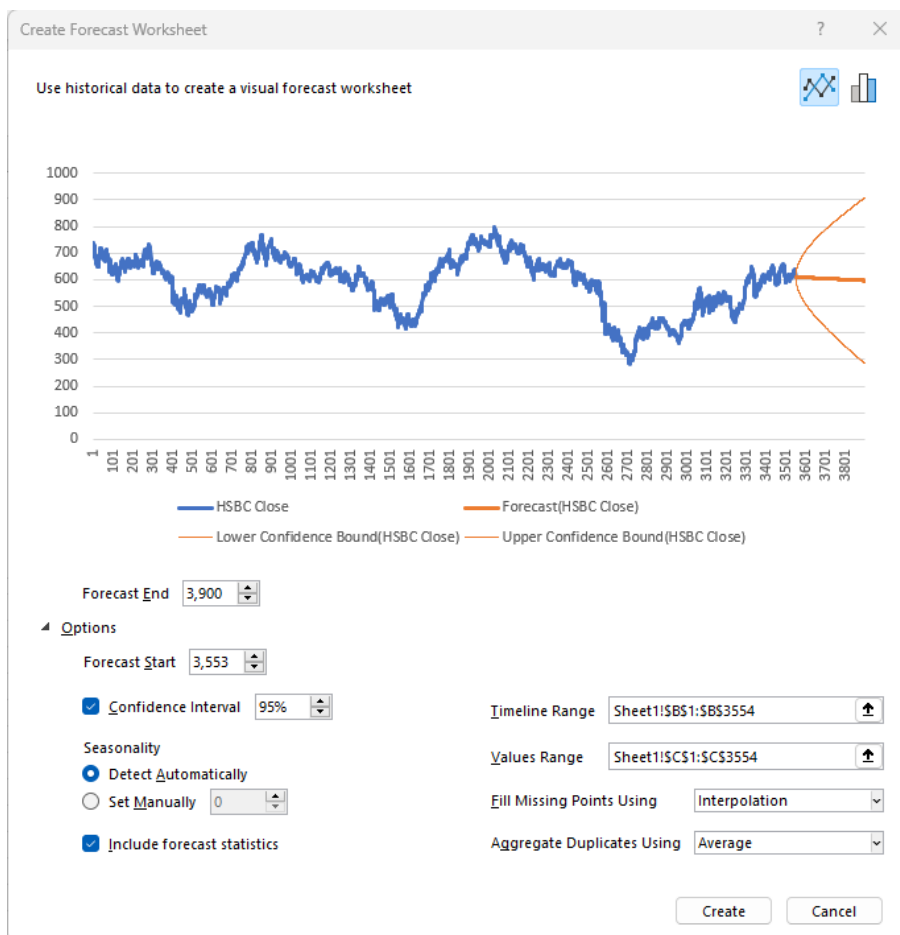


Fig. 7.6

As we can see from Fig. 7.6, the system could not automatically detect a seasonality and we effectively got a descending linear trend. Fig. 7.7 shows the full graph.

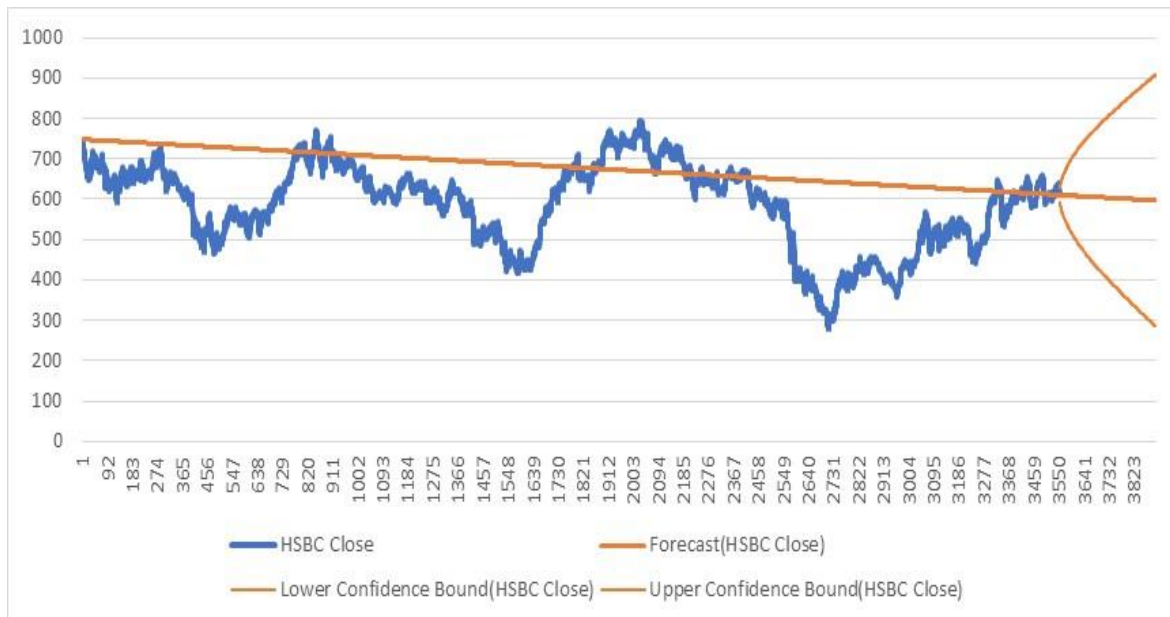


Fig 7.7

This is not what we were hoping to get, so let's try to use the function, as in cell J2 in Fig. 7.4 to detect seasonality. Unfortunately, this did not produce meaningful results. Here is an alternative. Take a look at the peaks of the fluctuating HSBC stocks in Fig. 7.5.

After you visually identified the peaks on the graph, search the dates in the spreadsheet when these peaks take place. You will find that the peaks took place on the following dates and that the number of days difference between them was as follows:

Peak	Number of days from peak to peak
06/01/2010	
20/05/2013	1230
16/01/2018	1702
17/10/2023	2100
Average	1677

If we take an average of these days' differences, we get 1,677 days as a potential seasonality (or more appropriately, the length of a cycle). We can plug this number manually into the seasonality box. When we did this, we got a graph instantly as shown in Fig. 7.8.

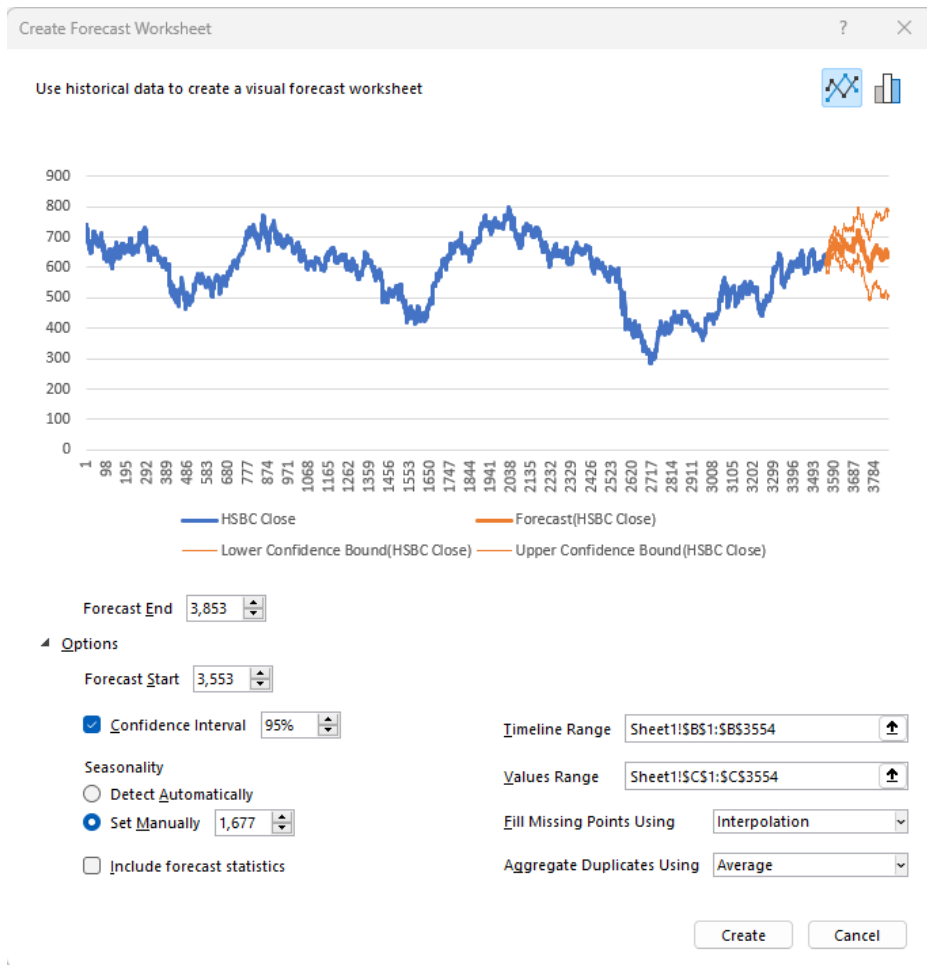


Fig. 7.8

In Fig. 7.9 we enlarged the graph and are showing how well this method fitted historical data, as well as the forecasts with the associated prediction interval.

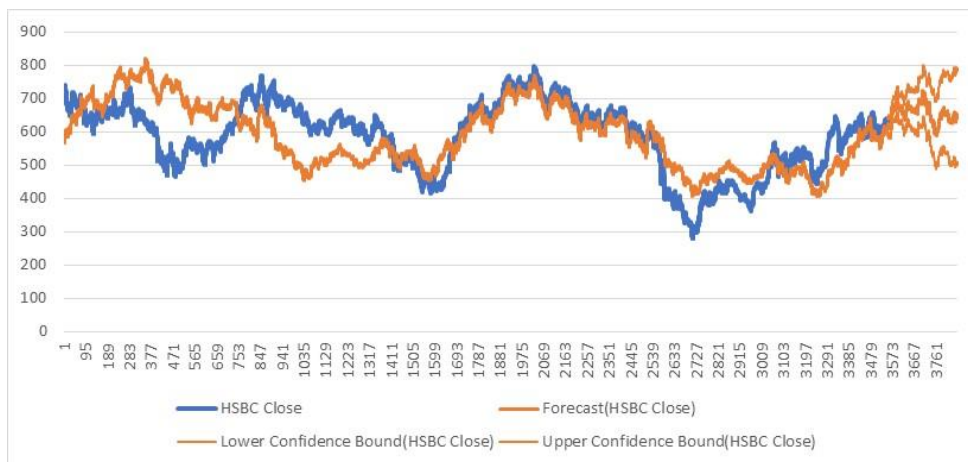


Fig. 7.9

When you look at Fig. 7.9 where we fitted the whole history and extrapolated it 300 days in the future, you see a very credible fit and forecasts.

So, this is how to use this method, but this is not the end of the story. Now you know how to use it, the question is: if we could have produced an even better forecast, and equally important, if we produce more forecasts, which one would be the best?

Before we do this, let's run a few quick simulations as this Excel routine enables us to partially see the results even before we click on the Create button.

We used the same HSBC series and the only change we made was to vary the length of seasonality.

Fig. 7.10 consists of four rows of graphs, two per row, that were shown the moment we changed the seasonality in the dialogues box. Starting from the top left, the seasonalities used were: 1100, 1230, 1500, 1550, 1600, 1700, 1800 and 2100.

As we can see some values result in downward-moving forecasts and a few (the last three) with an upward-moving forecast. Which one is the best one?





Fig. 7.10

To answer this question, We need to look at the error metrics that were given for all simulations in cells G1:H8. However, to make it a bit simpler, I only extracted five forecasts, i.e. those for the seasonality of 1230, 1550, 1677, 2100 and the one where the seasonality was automatically assigned by Excel (effectively a linear trend).

To pick the most suitable ETS model, we used the error metrics that Excel automatically produced (MASE, SMAPE, MAE and RMSE), but we also added a few more. They are MPE, RSE, R^2 and SE, and we have encountered some of them before. Fig. 7. 11 shows these error measurements.

	A	B	C	D	E	F
1		Auto	P=1230	P=1550	P=1677	P=2100
2	MASE	0.99	1.24	0.93	1.28	0.41
3	SMAPE	0.011482	0.015153	0.010901	0.015386	0.004251
4	MAE	5.92	7.36	5.51	7.60	2.43
5	RMSE	8.19	10.13	8.01	11.72	5.97
6	SE=	132.85	130.31	99.34	83.72	92.21
7	MPE=	-19.55032	-20.19276	4.22217	-1.64148	-10.64642
8	RSE=	30.23	26.87	17.63	15.63	16.25
9	RSQ=	0.15	0.47	0.39	0.42	0.60

Fig. 7.11

Before we analyse the data, let's just give you more background about one specific error metric (RSE), and remind you of R^2 , which in this instance we named RSQ.

The error metric in row 6 in Fig. 7.11 is the standard prediction error (SE). As we know this is a goodness of fit measure that we used to define the prediction interval around our forecasts. To remind you, the width of the prediction interval will depend on the size of this standard error, as well as on the level of confidence we want to use (typically 95%, but could be any other level).

In the case of Auto forecast, for example, the SE is 132.85. The HSBC values fluctuate roughly around 580, and this SE is expressed in the same units as the value of the HSBC stock (which is GBP). This means that ± 132.85 around any value of the historical stock is an interval around which a majority of the stocks will fluctuate (plus the uncertainty level defined by the level of confidence).

This SE could be expressed in relative terms, where we calculate the average percentage around every historical value. The percentage defines the "zone" around the actual value and the fitted values are expected to be inside this zone. In other words, it is another goodness of fit measure, but expressed in percentages. For example, for the Auto solution from Fig. 7.11, we have RSE = 30.23, which means that forecasts deviate from the actual values on average by 30.23% of the actual values. In other words, the fitted line creates a "band" or a "zone" around the actual values, and this zone is in this case on average 30.23% wide. This is not a good result. Statisticians would usually reject any fit that shows RSE above 30.

The last error metric that we wanted to remind you of is RSQ, or R^2 , which is the coefficient of determination we met before. However, this time we used it somewhat differently, which is the reason we used the abbreviation RSQ, rather than R^2 . If you remember we used this metric to measure how well are the variations in y explained by the variations in x . Our actual stock data were y and just sequential numbers from 1 to n , representing the time periods, were x .

In this case, we are still using actual stock data as y , but instead of x , we use \hat{y} which represents the fitted values. In other words, how much are the variations in fitted values explained by the variations in actual values (just to clarify, mathematically it does not matter if we use one or the other for \hat{y} or y in this equation, we get the same result).

When you see RSQ = 0.15 for the Auto fitted line from Fig. 7.11, this means that only 15% of all variations in the fitted line are explained by the variations in actual values. Not a good result.

Anyway, we can go back to Fig. 7.11 and analyse the data we got by comparing five different fits to our historical HSBC data. I highlighted the best values in green. We can see that the competition is

only between the fit that uses 1,677 days for the seasonality (periodicity is a better word here) and the one that uses 2,100 days. Let's analyse it.

For the periodicity of 2,100 days, we have the lowest MASE (implies highest accuracy), lowest SMAPE (highest fit), smallest MAE and RMSE (also measures of accuracy) and the highest RSQ (highest dependence of the fit and actual data). Out of 8 different metrics, we have 5 of them scoring the best scores.

For the periodicity of 1,677, we have the lowest MPE (a measure of accuracy) and the lowest RSE and SE (measures of goodness of fit). Here we have the three out of eight highest scores. If we just take the numerical advantage, then the 2,100-day periodicity forecast is superior to the one for 1,677 days. But, is it?

This is the point where predictions wander away from the world of science into the world of art. Although the numbers above suggest that we should give preference to the 2,100-day periodicity forecasts, this might not be the right conclusion. This is where other factors begin to play.

Let's first take a look at how both fits and forecasts look like. Fig. 7.12 shows the chart with the actual values, the two fits (1,677 and 2,100 days periodicity) and the two forecasts.

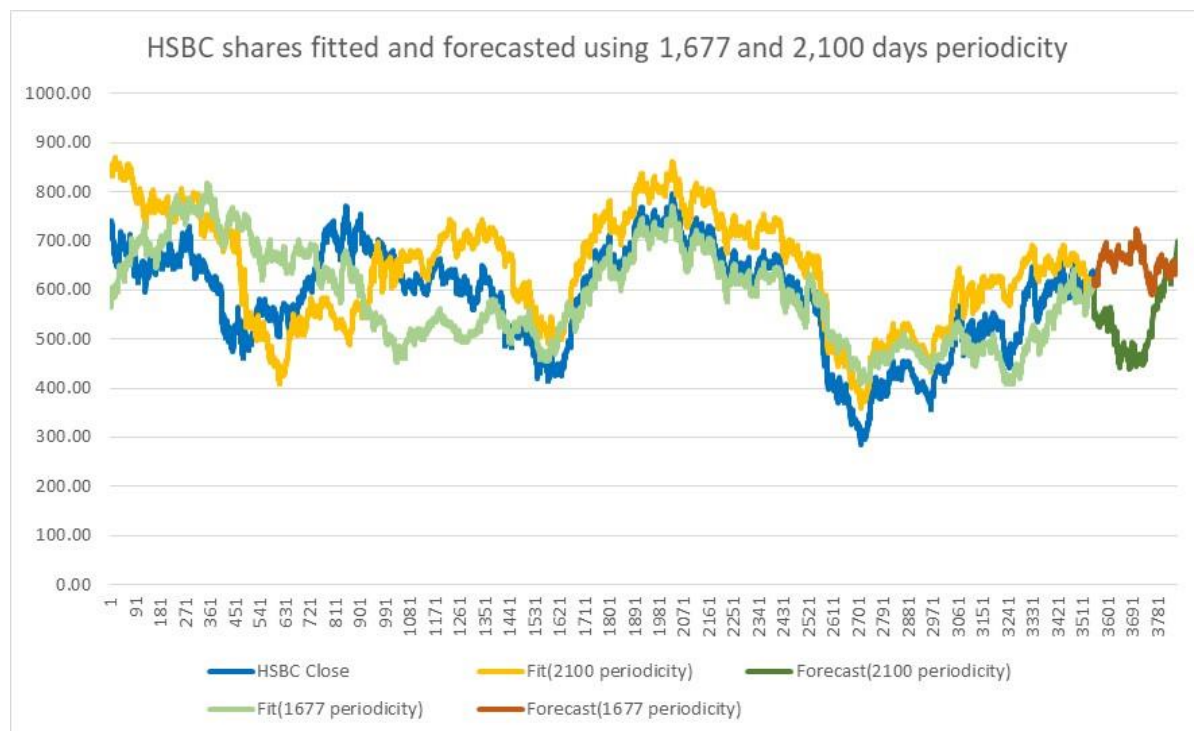


Fig. 7.12

The 2,100-day periodicity forecast implies that HSBC closing shares will continue to fall for another 130 trading days reaching the level of below GBP 450. Eventually, they will start to rise, reaching the value of almost GBP 700 after 300 trading days from 29 Dec 2023 (which will be somewhere around the 20th of Feb in 2025, i.e. just over a year from 29 Dec 2023). See Fig. 7.13 showing these forecasts.

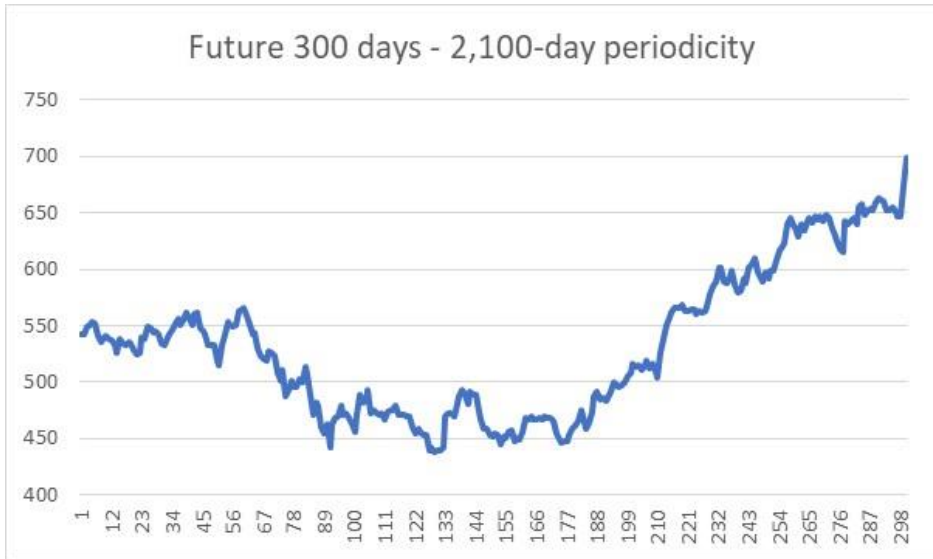


Fig. 7.13

The 1,677-day periodicity forecast anticipates more or less horizontal future moves of the closing share values. Over the next 300 trading days, the value will only once marginally drop below GBP 600 and after 300 trading days, the share price will be marginally below GBP 650 (642.9, in fact). See Fig. 7.14 showing these forecasts.

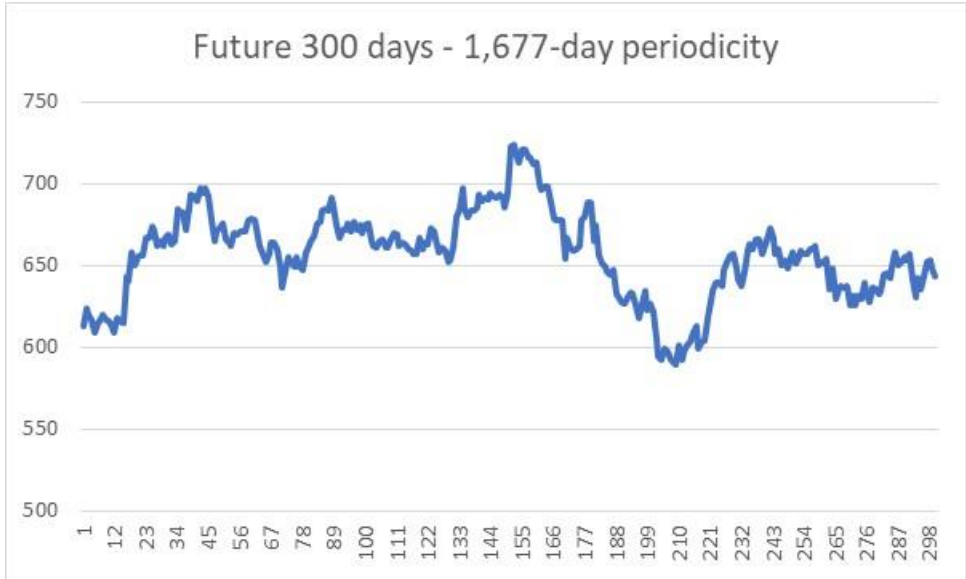


Fig. 7.14

To make it a bit more visible, we took just the last 300 trading days of our time series, as well as the future 300 forecasts. Fig. 7.15 shows this.

As we can see, both forecasts will after 300 days end up close to one another, but in between will take two different routes. The 1,677 periodicity forecasts will continue with a gentle growth and the 2,100 periodicity one will first drop and then recover to a similar number after 300 days. Which one is correct?

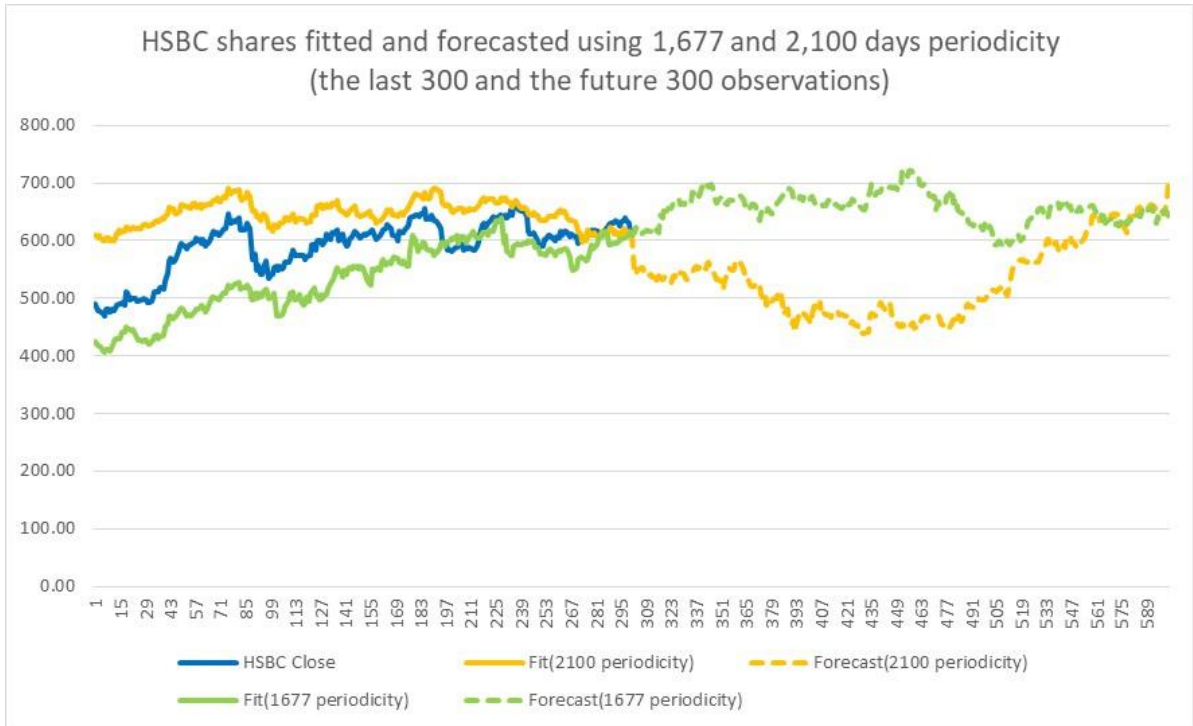


Fig. 7.15

To be frank, I do not know which one is correct! If you are reading this 300 days after we produced these forecasts, you can figure out which one was better!

However, I can tell you what happened around 2 Feb 2024, which is the time I last revised this chapter. See Fig. 7.16.

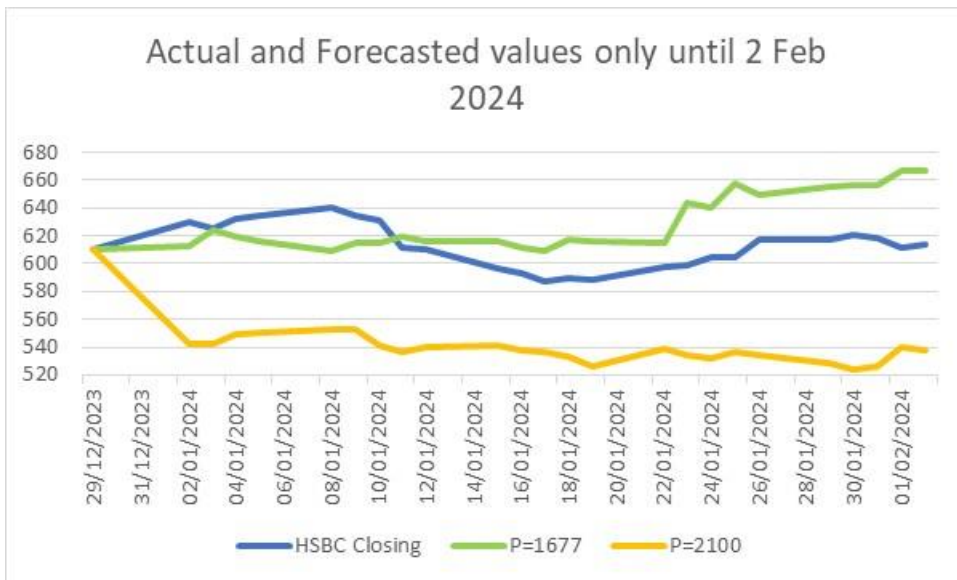


Fig. 7.16

According to these first 24 trading days, the 1,677-day periodicity forecasts seem to be initially tracking better the actual values than the 2,100-day periodicity forecasts, though it started to diverge after the first 8 days. We'll see what happens next.

This also brings us to a difference between the short-run vs. the long-run forecasts. So far we only concentrated on the long-run forecasts that are typically fitted with a trend line of some sort, or in this case, by the Excel ETS method.

Could these methods be also used for the short-term forecasts, or do we have any alternative methods? Again, we'll stick to what Excel gives us and tackle the short-term forecasts in the next chapter.

9. Predicting a turning point in stocks and shares movements

Imagine the following scenario, you are monitoring the daily movements of your stocks and shares and you thought that it would be useful if you could recognise that the direction has changed. In other words, if the stocks were following a downward trend, you would like to know that they have suddenly started moving upwards. The same applies to the opposite scenario where after an upward trend, you suspect a downward trend will follow.

From the description above, it is clear that a turning point in stocks, sometimes also known as an inflection point or pivot point, is a moment when a significant change in the direction of the stock movements takes place.

The interesting part is that if we take a look at historical stock movements, we can reasonably well recognise the points at which these changes take place, i.e. we can clearly see the turning points.

Having said that, we know that often the values change direction, but then for a while, they keep on oscillating up and/or down, before definitely setting a new upward or downward trend. So, the trick is to predict a genuine turning point, as opposed to just one of these irregular ups and downs.

An important disclaimer here is that when we use the word prediction, we do not mean predicting future turning points. This would be a bridge too far. What we mean is predicting, in this moment of time, that the turning point is actually taking place. You check your stocks, and you conclude that the stocks are likely to start a downward or upward run. This enables you to take action, which is to sell, buy, or whatever else is on your mind.

Before we show how to do it, just one word of caution. The challenge here is the question: What exactly constitutes a turning point? To illustrate this, let's look at two identical graphs below in Fig. 9.1.



Fig. 9.1

We took GSK stock closing values between 7 Mar 2023 and 7 Sep 2023 and just drew some straight lines that potentially represent the trend that is being followed over a particular period of time. The graph on the left uses only three lines and generates only two turning points. The graph on the right shows four turning points as we have more local trends.

The question is: which graph is more correct? The answer is: Both are equally correct! Let's explain and use an analogy.

You know how when you look at a weekly graph of any stock movements, you see a lot of "wiggles" and ups and downs. However, when you see the same data, but included in a graph that covers a

longer period of time (a month or a year, for example), then these weekly ups and downs become less significant. The same is true with this example.

A local trend is defined by you and what you consider to be relevant. For a person who constructed the graph on the left, he or she is only interested in major shifts in direction. The person on the right is more interested in finer changes because the strategy that he follows depends on identifying these subtler changes in direction.

To summarise, we need to add a level of “sensitivity” in our charts and, depending on what we want to choose, we will identify either a fewer or greater number of turning points.

Let’s show how to build a model in Excel to help us identify that the turning point is about to happen, or has just happened. We’ll use GSK daily closing values between 1 Mar 2023 and 28 Apr 2023. In total, only 41 observations. Fig. 9.2 shows the chart.



Fig. 9.2

The chart in Fig. 9.2 is produced from the data in columns A and B in Fig. 9.3. We are hiding rows 16:33 just to get a clearer picture of what is happening in the spreadsheet. Let’s see how is our model built.

Columns A and B come directly from the stock exchange. Column C contains the moving averages of column B. If you take a look at cell C7, you will see the function =AVERAGE(B2:B6). Cell C8 contains =AVERAGE(B3:B7), etc. Every value in column C is a rolling average of the previous five values from column B. These rolling averages are called moving averages and we already used them in the previous chapter.

Before we proceed to columns D and E, pay attention to cells I1:J1. They say that L=1. This is the number of standard deviations that we will use to calculate columns D and E. If we wanted to use less sensitivity for our data, we could have used L=2 (just replace the value of 1 to 2 in cell J1).

Cells D7 and E7 have the following functions:

$$D7 = C7 + \$J\$1 * STDEV.S(B2:B6)$$

$$E7 = C7 - \$J\$1 * STDEV.S(B2:B6)$$

D7 is the beginning of the Upper Limit (UL) and E7 is the first cell in the Lower Limit (LL) column.

	A	B	C	D	E	F	G	H	I	J
1	Date	GSK	5MA	UL	LL	IF	Clean	Turn	L =	1
2	01/03/2023	1422.00								
3	02/03/2023	1438.80								
4	03/03/2023	1442.20								
5	06/03/2023	1439.40								
6	07/03/2023	1441.80								
7	08/03/2023	1422.40	1437	1445.27	1428.41	Change DOWN	Change DOWN	1,422.40		
8	09/03/2023	1412.20	1437	1445.17	1428.67	Change DOWN				
9	10/03/2023	1398.60	1432	1445.19	1418.01	Change DOWN				
10	13/03/2023	1376.60	1423	1441.15	1404.61	Change DOWN				
11	14/03/2023	1380.40	1410	1434.89	1385.75	Change DOWN				
12	15/03/2023	1381.40	1398	1417.82	1378.26	Change DOWN				
13	16/03/2023	1387.00	1390	1404.94	1374.74	Change DOWN				
14	17/03/2023	1400.80	1385	1393.37	1376.23	Change UP	Change UP	1,400.80		
15	20/03/2023	1414.00	1385	1394.70	1375.78	Change UP				
34	18/04/2023	1484.00	1514	1519.46	1507.82	Change DOWN	Change DOWN	1,484.00		
35	19/04/2023	1485.80	1509	1523.83	1494.49	Change DOWN				
36	20/04/2023	1470.80	1504	1521.49	1486.11	Change DOWN				
37	21/04/2023	1471.20	1493	1512.60	1474.36	Change DOWN				
38	24/04/2023	1474.00	1485	1501.39	1468.05	Change DOWN				
39	25/04/2023	1500.20	1477	1484.36	1469.96	Change UP	Change UP	1,500.20		
40	26/04/2023	1442.20	1480	1493.04	1467.76	Change DOWN	Change DOWN	1,442.20		
41	27/04/2023	1437.00	1472	1492.23	1451.13	Change DOWN				
42	28/04/2023	1441.00	1465	1490.72	1439.12	Change DOWN				
43			=AVERAGE(B37:B41)							
44				=AVERAGE(B37:B41)+\$J\$1*STDEV.S(B37:B41)						
45				=AVERAGE(B37:B41)-\$J\$1*STDEV.S(B37:B41)						
46						=IF(B42>D42,"Change UP",IF(B42<E42,"Change DOWN",F41))				
47							=IF(F42=F41,"",F42)			
48								=IF(G42<>"",B42,"")		

Fig. 9.3

Why do we have the Upper and Lower Limits and what do the functions say?

The strategy is as follows: We will compare every current value of the stocks with its previous five-day average. If it is significantly different, either up or down, then potentially we have a deviation from the existing trend and the occurrence of the turning point.

The Upper and Lower Limits serve to signal to us if this deviation from the current trend is “significant” and we define this word significant as deviation beyond \pm one standard deviation (or two, if you want less sensitive data).

The functions from cells D7 and E7 are just Excel translations of the equation:

$$\bar{y}_5 \pm L \sigma_5 \quad (9.1)$$

Where,

\bar{y}_5 = The mean value of all the previous five actual prices (a moving average)

σ_5 = Standard deviation of all the previous five observations

L = Number of standard deviations

Equation (9.1) is a variation on the concept of prediction interval from Chapter 5, so you will be familiar with this notion.

The Upper Limit is the expression where we add $L \times \sigma_5$ to \bar{y}_5 and the Lower Limit is where we subtract $L \times \sigma_5$ from \bar{y}_5 . The value of L comes from cell J1 in Fig 9.3.

From the functions in cells D7 and E7 (as well as all the subsequent cells in the rest of columns D and E), you can see that just as we are using moving averages for the past five values, we are also using standard deviations that apply to only these five past observations. So, our Upper and Lower Limits are dynamic.

In cell F7 we have a function: `=IF(B7>D7,"Change UP",IF(B7<E7,"Change DOWN",F6))`. This formula is copied down the column and it simply says that if the actual value (from column B) is greater than the Upper Limit based on the past 5-day average (column D) then we should see the phrase "Change UP" in this cell. If the actual value (from column B) is smaller than the Lower Limit based on the past 5-day average (column E) then we should see the phrase "Change DOWN" in this cell.

As you can see column F is full of repeated phrases, so to make it a bit clearer, we created column G, where in cell G7 we are using a function: `=IF(F7=F6,"",F7)`. In other words, let's eliminate the repetitions of the phrase and clearly label only the turning points.

Next to it, a cell in H7 we have a function: `=IF(G7<>"",B7,"")`. This simply says that only where there are the phrase indicating the turning point, we will copy the value from column B into column H. This will help us to build a graph of all the turning points. We did this in Fig. 9.4.

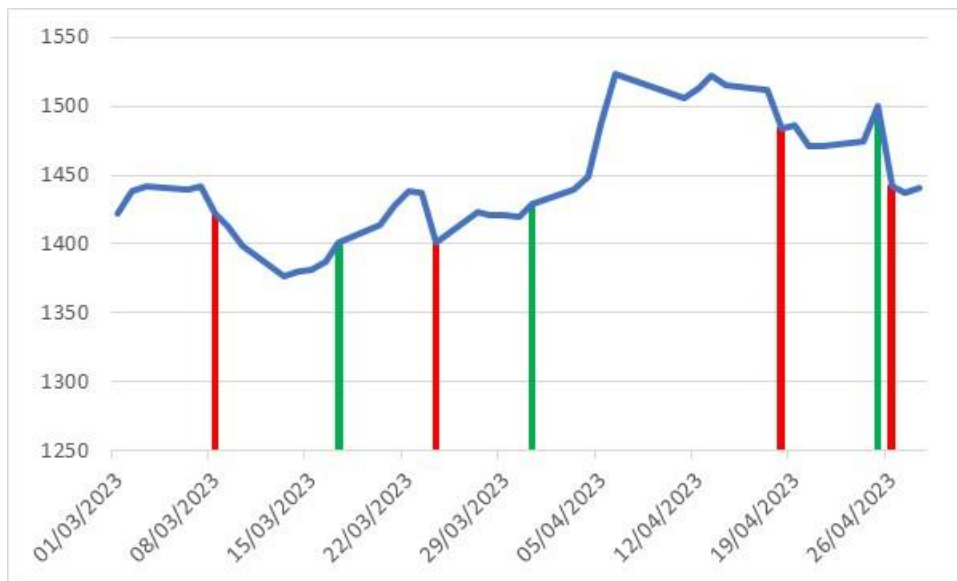


Fig. 9.4

Just to make it a bit more visible, we manually coloured the vertical lines that represent the turning points. The red line indicates a downward change and green represents an upward change.

How good are these historical fits of the turning points? Well, not bad at all. We can see that we are lagging behind the actual turning point by 1-3 days, which is OK. First of all, this is caused because we are using moving averages as a reference point. Secondly, you do not want to overreact as not every drop or jump is the beginning of the actual turning point.

We can take a much longer time series, again GSK, but between 22 Feb 2023 and 14 Nov 2023 (185 observations) and calculate the turning points for $L=1$ and $L=2$. We have done this and Fig. 8.5 shows the results. The graph on the left ($L=1$) shows a larger number of turning points than the one on the right ($L=2$). Clearly, the Upper and Lower Limits for $L=1$ are narrower, so we get more turning points.

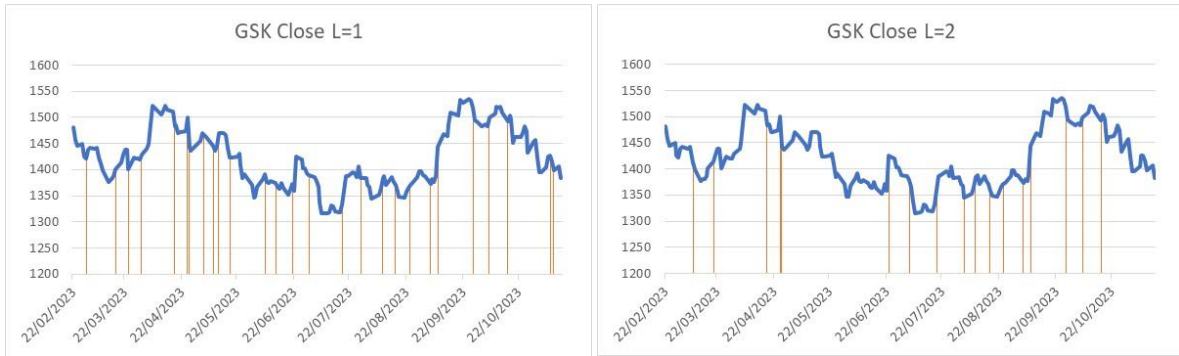


Fig. 9.5

Regarding what we’ve just shown, we want to make a point and recommend that one should not take a time series of stock values that is longer than 90 observations. It just does not add any value. You are really not interested in historical turning points. You are interested in detecting one happening in real time. For this purpose, thirty to sixty days is probably more than adequate.

Just to use several more examples, we picked four other stocks and produced similar charts that apply to the last 60 days, where the “current” date was 29 Dec 2023, i.e. the last trading day in 2023 (see Fig. 9.6).



Fig. 9.6

In all the cases, the turning points were delayed by a couple of days on average. We explained the reasons already, and this is perfectly OK. From Fig. 9.6 we can see that GOOG and OMG seem to be on the brink of a turning point. In both cases, the movement is about to change to a downward trend. Well, by the time you read this, we’ll know what happened.

As a closing remark, we might ask ourselves just one more interesting question, which is: when the turning point happens, how many days before the trend changes again?

Believe it or not, you'll struggle to find an answer to this question. However, I did some number crunching, and although what I am about to say is not necessarily 100% representative, it is probably very close to the truth.

I took 37 different shares quoted on the NYSE, NASDAQ, LSE and Frankfurt Exchange and examined approximately the last 3,500 trading days (unless the shares were not quoted on the exchange for this period of time). I did a simple analysis and checked daily returns, so if the closing value went up, it was marked as U, and if it went down it was marked as D. I looked at these runs of U or D changes.

Here is what I discovered:

1. An average run is 7-8 days, with U runs lasting on average 8 trading days and D runs 7 trading days
2. The maximum number the shares spent in the U run is on average 39 days, whilst the maximum for the D run is 30 days, based on the portfolio of 37 shares I analysed
3. The longer the historical dataset, the higher the ratio of the U runs vs. D runs
4. For longer datasets, the ratio of U vs. D is approximately 1.15:1 and for shorter datasets, the ratio of U vs. D is approximately 1:0.97 or 1:0.99
5. The greater the sensitivity for detecting the turning points, the more observations are found in the U runs. By inference, this implies the higher ratio of U vs. D

These are just some of the findings I unearthed as digging through the share movements. Hopefully, this makes sense and can help you decide how to handle your shares once you identified that you are potentially on the cusp of a turning point.