Time series and Exponential smoothing

Time series [Schumway]

A *stochastic process* is a set of random variables $\{y_t, t \in I\}$.

The result of an observation of these variables is a *realization* of the stochastic process, also called *time series*.
It is customary to also refer to the process giving rise to a time series as a time series.

If the indexing set I is countable (uncountable), then the time series is said to be *discrete* (*continuous*).

Forecasting methods and models [Hyndman]

Forecasting method: algorithm that gives a prediction of the value of the series at a certain future time.

Forecasting model: stochastic data generating algorithm, producing a prediction of the probability distribution at a future time period.
A point forecast can be obtained by taking the mean of this probability distribution.

Exponential smoothing forecasting methods [Hyndman, Athanasopoulos]

It is common to try to split a time series into three components: the *trend* (or level) (T), *seasonal* (S) and *error* (E) components.
These components can be combined in different ways.
For example by simply adding them, so that y = T + S + E, or by assuming multiplicative seasonality but additive error, so that y = T x S + E.

The exponential smoothing methods ignore the error component, and have corresponding models (more specifically state space models). We will only discuss the methods, not the models.

Exponential smoothing was introduced in the late 1950's by Brown, Holt and Winters.
Roughly speaking, an exponential smoothing method forecasts using a weighted average whose weights decrease exponentially as the corresponding observation gets older.
There are different exponential smoothing methods, which we will briefly discuss.

Simple exponential smoothing

Use: forecasting using data with no trend and no seasonality.

To understand simple exponential smoothing it helps to first describe the *average forecasting method*, which is one of the simplest forecasting methods.
In this case, the forecast is simply the average of the observations:
<p align="center">

</p>
where |T in the subscript means we are using the T first datapoints for the forecast.
This is clearly too simple for most cases.

Simple exponential smoothing (SES) is similar to the average method, but the average is weighted, with the weights decreasing exponentially "with age".
More precisely:

To fit the time series + forecast the next point:
<p align="center">

</p>
where $0 < \alpha < 1$ is the so-called *smoothing parameter*.
To forecast further in the future one just uses the last predicted value:

Remarks:

- Notice that the sum of all the weights is
  <p align="center">

  </p>
  as it should be in a weighted average.

- SES simply forecasts a horizontal straight line, thus not accounting for seasonality or trend.

- A good value for α is usually found using least squares (minimizing the sum of squared errors (SSE)).

- The way we defined SES is fairly natural.
  It turns out that we can formulate SES in another form - called the *component form* - which will be useful for future generalization of this method.
  It consists of two equations:
  <p align="center">

  </p>

Holt's linear trend method

Use: forecasting using data with an approximately constant trend but without seasonality.

Holt's linear trend method is an extension of SES which separates T in two components: the level l and the trend (slope) b; it modifies the forecast and smoothing equations and introduces a new equation for the new component b as follows:
<p align="center">

</p>
where $0 < \alpha < 1$ is the smoothing parameter for the level and $0 < \beta^* < 1$ is the smoothing parameter for the trend.

Remarks:

- The level l_t at time t is still (like in SES) a weighted average between y_t and the estimate of y_t (using the data points up to time t-1).

- The trend b_t at time t is a weighted average between the natural estimate $l_t - l_{t-1}$ of the slope/trend and the previous trend b_{t-1} .
- The forecast is no longer flat! Instead, it is a linear function of h.

Holt's damped trend methods

Use: forecasting using data with a decreasing trend but without seasonality.

It is clear that the Holt's linear trend method tends to overshoot in many cases for long forecast horizons (h substancially greater than 1). One can try to solve this by introducing damping of the trend such that the level does not increase indefinitely.

Holt's damped trend method is similar to Holt's linear trend method, but includes a *damping parameter* $0 < \phi < 1$ as follows:
<p align="center">

</p>

Remarks:

- It is easy to see how b_t and l_t are damped. To see that the forecasts of y_t get flatened (capped slope) for high h, just notice that:
  <p align="center">

  </p>
  So *e.g.* for $\phi = 0.9$ the forecast approaches the line $l_t + 9b_t$ (its slope never surpasses $9b_t$).

Holt-Winters' additive seasonal method

Use: forecasting using data with an approximately constant trend and with seasonality, and the seasonal variations are approximately constant accross time.

Note: If the seasonal variations are approximately proportional to the level, then one uses the so called multiplicative seasonal method.

Holt-Winters' additive seasonal method is similar to Holt's linear trend method, but includes a ** $0 < \phi < 1$ as follows:
Holt-Winters' additive seasonal method is an extension of Holt's linear trend method which adds the seasonal component S introducing a new equation for the new component s_t as follows:
<p align="center">

</p>
where m is the *seasonality frequency* (number of data points per season) and k = floor((h-1)/m).

Remarks:

- The level equation simply subtracts the seasonal part from y to get the level estimate. The trend equation of course needs no correction.
- The seasonal equation is a weighted average of the seasonal component estimate (y_t - l_t) and the seasonal component s_{t-m} from the last period.
- The forecast equation contains $s_{t+h-m(k+1)}$ so that it always used the last fitted

seasonality (*i.e.* it does not use the forecasted seasonality for new forecasts!). *e.g.* if h = 3m + n where n is smaller than m, then k=3 and the seasonal component added to the forecast is $s_{t+h-4m} = s_{t-m+n}$.

Table with all Exponential Smoothing methods

We can introduce damping in seasonal methods and assume multiplicative seasonality instead of additive seasonality.
All these options together give us a list of all the exponential smoothing methods:

	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N, N)	(N, A)	(N, M)
A (Additive)	(A, N)	(A, A)	(A, M)
Ad (Additive damped)	(Ad, N)	(Ad, A)	(Ad, M)

	Seasonality		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N, N)	(N, A)	(N, M)
Trend: A (Additive)	(A, N)	(A, A)	(A, M)
Ad (Additive damped)	(Ad, N)	(Ad, A)	(Ad, M)

TODO:

- Add plots.

References:

- [Schumway]:
- [Hyndman]:
- [Athanasopoulos]: