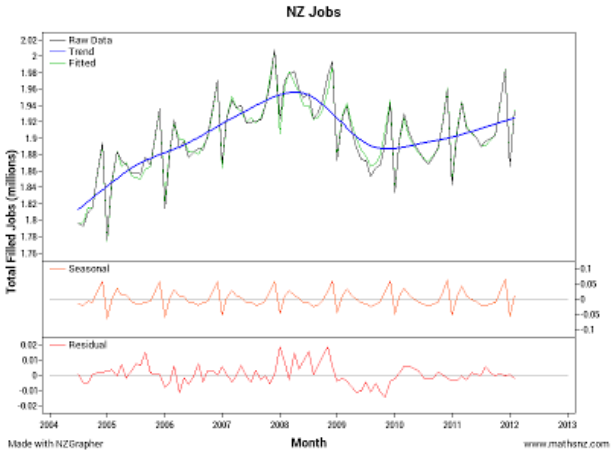
What is in the 'black box' of NZGrapher and iNZight's time series module?

The short answer is Seasonal Trend LOESS Decomposition for historical data and Holt-Winters for forecasting... but what does this actually mean? The data used in this example is the number of people in employment in New Zealand. The data is from the National Employment Indicator series sourced from Statistics New Zealand[1].

Seasonal Trend LOESS (STL)

STL is a method of decomposing historical data to analyse time series based on a long term trend and a seasonal component. It was formulated by Robert B. Cleveland and others, and published in the Journal of Official Statistics in 1990. You can access the full article[2].

As a brief summary, it runs through the following loop (for NZGrapher it does this twice), starting with a completely flat trend.



1. Subtracting the trend from the raw data.
2. Getting all the values for each season (eg: all the January values) and smoothing them using LOESS* extended by one year before and after and then combining these back into one series.
3. A moving mean with a length equal to the number of time periods is fitted to the series from step 2 twice to centre it, followed by a LOESS smoothing.
4. The seasonal values are calculated by subtracting the series created in 3 from the series created in 2 and then averaged.
5. A deseasonalised series is then created by subtracting the seasonal values (from 4) from the raw data.
6. This deseasonalised series is then smoothed using LOESS and this gives the trend values that are used for the next loop through.

Once the loop is finished going through the required number of times, the deseasonalised series from step 6 forms the trend (blue line), and the seasonal values from step 4 form the seasonal value (orange line). The residual values are calculated by subtracting the trend and the seasonal values from the raw data (red line). The fitted values (green line) is the trend plus the seasonal.

The STL method can be set to use either an additive or a multiplicative model. We'd described the process for the additive model above. The process for the multiplicative model is similar. The STL method aims to decompose the raw data (the actual, for some people) into three components:

Raw = Trend + Seasonal + Remainder or Raw = Trend * Seasonal * Remainder

* LOESS can be thought to stand for LOcal regrESSion and works by selecting a point, getting the closest n points to it, fitting a weighted regression line through them, and working out what value this has at the selected point. The n is a parameter that can change depending on the data.

Holt Winters

The Holt-Winters model requires a number of parameter values to be chosen, the first few are the initial values for the trend, the gradient of the trend and the seasonal effects.

To start off with NZGrapher creates a centered moving mean for the first two years of data. Once it has done this it fits a linear trend to it. The the y-intercept is the initial value of the trend, and the gradient of

---

[1] The NEI was a series produced by Statistics New Zealand and covers filled jobs where employees were paid wages or salaries in the calendar month, by an employer who filed an EMS return. This series was discontinued in 2012. More details.

[2] http://cs.wellesley.edu/~cs315/Papers/stl%20statistical%20model.pdf

this line is the initial value for the gradient for the long term trend. The initial values for the seasonal effects are the difference between the first year values and the value of the linear trend line.

The general formula for the Holt-Winters model is:

$$\hat{y}_{T+\tau} = a_T + \tau\ b_T + s_T$$

Where a_T is the current trend level, b_T is the current gradient of the trend and s_T is the appropriate seasonal effect.

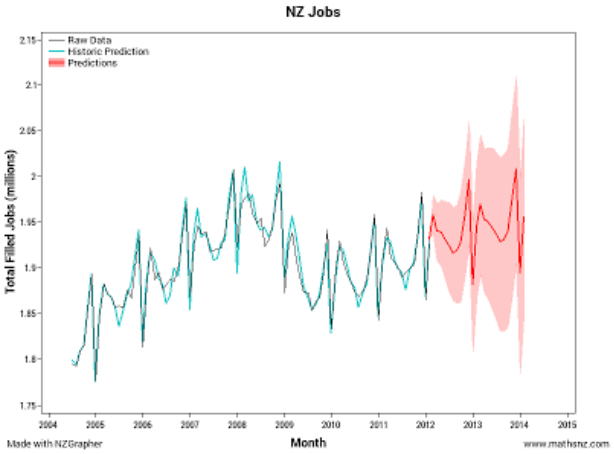
The Holt winters uses the following formulas to work out the values of a_T, b_T and s_T as it goes through:

$$a_t = \alpha(Y_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$

$$s_t = \gamma(Y_t - a_t) + (1-\gamma)s_{t-p}$$

These are what is known as exponential smoothing formula, and as you can see they have three extra parameters: a, β and γ.



These are worked out by making a prediction for each step of the time series and trying to minimise the squared error for this prediction that is one step ahead. NZGrapher goes through and tries a number of different combinations, refining the values of a, β and γ until it has a model that fits as best as it can. Once it has done this it draws in the blue and green lines on the time series recomposition graph.

The difference between the aqua and black lines, or the error between what the model was predicting would happen and what actually happened is what the software tries to minimise. Because of the nature of exponential smoothing, the most recent values always have the greatest impact on the forecasts that we will make.

The software uses the final values of a_T, b_T and s_T to make the main predictions (shown by the solid red line on the software, and for the pink bands of expected values it does a bootstrap process 1000 times of what could happen from normal variation from what is seen in the data, and takes the middle 950 of them to give the most likely outcomes. This is why each time the forecast output is produced the values change slightly.

This method uses all the historical information (i.e. all the points in the series) to estimate: the current level of the series; the current slope of the series; and the current value of the seasonal factors. It uses these values to make these predictions. The method uses exponential weighting to ensure the most recent values have the largest weight, and the weights decrease with age. The parameters a, β and γ determine how fast the weights drop off with age.

Hopefully this sheds some light on what is going on inside the 'magic black box' of NZGrapher and iNZight, and if you have any questions about this don't hesitate to ask. This is not something that students are expected to know or understand, but it is useful to be able to point in this direction for students who are wanting to really understand what is going on.

Jake Wills
With input from the New Zealand Statisitical Association Education Committee
jwills@mathsnz.com