Aston Business School

Time series analysis

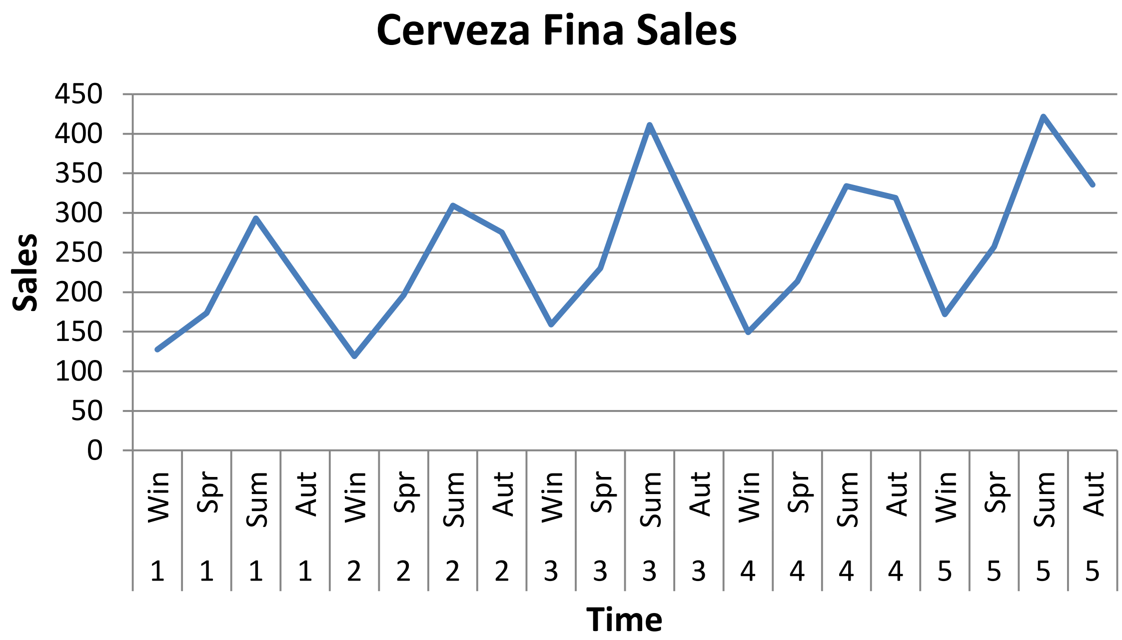
BN2255 – Business Analytics in Practice

Case study: Cerveza Fina

Year	Season	Sales
1	Win	127
1	Spr	174
1	Sum	293
1	Aut	205
2	Win	119
2	Spr	196
2	Sum	309
2	Aut	276
3	Win	159
3	Spr	230
3	Sum	411
3	Aut	279
4	Win	149
4	Spr	214
4	Sum	334
4	Aut	319
5	Win	172
5	Spr	257
5	Sum	422
5	Aut	336

- Cerveza Fina, a beer brewery operating in Spain, wants to get a better understanding of its sales patterns. For this reason, it has provided you with the following historic (past) data. The data cover a period of 5 years and present the sales of the company in each quarter. Use the provided data to generate a sales forecast that covers the next two years

Non-stationary data



- When our data reveals clear patterns and non-stable behaviour, MA-based forecasts are no longer appropriate
  - Average values of past performance are not longer a reliable indicator of the future
  - Does not mean that we can no longer attempt to forecast!
  - Instead, try a time-series model

Time-series models

- How to forecast a variable with complex patterns?
  - Try to isolate the source of these patterns
    - Break down the patterns into their component parts
- Three main components of time-series models
  - The trend – underlying, long-term movement in data series
  - Seasonality – repeated patterns that are somehow related with time
  - Randomness – random, unpredictable changes in our data
- Other components can also be included, such as cyclicality, reversion to the mean, random shocks etc
  - But these require more complicated statistical models
    - Out of scope for this module

$D = T + S + R$

Where: D = data series, T= trend, S = seasonality, R = random variation

Formulating the model

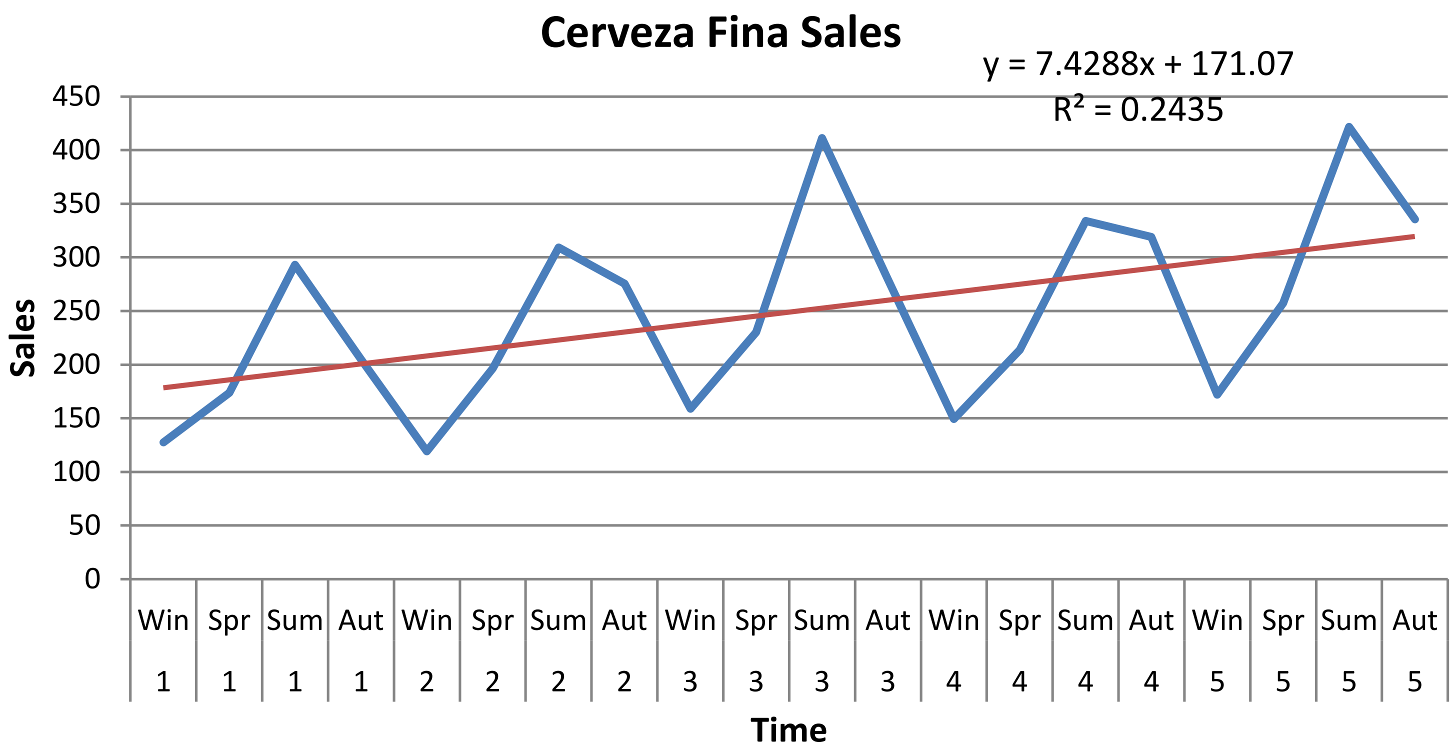
$D = T + S + R$

where:
D = data series, and
T= trend,
S = seasonality,
R = random variation

- What about the random variation component (R)?
  - since it is random, it is by definition unpredictable
  - we can measure it residually to get an idea of the accuracy of our model
    - sometimes even the best of models cannot predict the future!

- Data series is series of the variable we are interested in forecasting
- Trend is the underlying movement of the variable
  - Numerous ways to estimate trend
    - For this module, we focus on regression methods
- Seasonality represents the patterns in the data that are repeated over time
  - eg. ice-cream sales over the year
    - Can be estimated either residually or directly through regression

Deriving the trend



- Each session is a discrete time-period
  - Win-1 = 1, Spr -1 = 2, Sum-1 = 3, Aut-1 = 4, Win-2 = 5 ....
  - Trend Forecast for period 21, ie Win-6, is
    - 7.43*(21)+171.07 = 327.1
  - Note the coefficient of determination – R^2 = 24.4%
    - not a very accurate model

Estimating seasonality

- Main reason why the previous model had such a low R^2 was the substantial variation around the estimated trend line
- It is easy to see however that this variation is not totally random
  - actual sales are always lower than predicted sales in the winter and higher in the summer
  - strong evidence of seasonality
- Two ways to measure the effect
  - Residually: **S** = **D** – **T**
    - T is the trend forecast (backcast) of the simple time-series regression model
      - Good way to get a quick and easy estimate of the magnitude of the effects, but simplistic
  - Directly through the time-series regression model
    - Seasonal indicators are introduced directly to the model using dummy variables – see the week's case study for an example
      - Likely to result in more accurate estimates for both the seasonality effects and the trend!

Putting it together

	Coefficients	Standard Error	t Stat	P-value
Intercept	95.23	15.03	6.34	0.00
t	5.56	1.04	5.35	0.00
W	0.00	0.00	N/A	N/A
SP	63.39	16.67	3.80	0.00
SU	197.43	16.77	11.78	0.00
AU	121.01	16.93	7.15	0.00

- Adj R^2 = 84%
- Forecast for period 22 (Spring of year 6) using the model adjusted for seasonality:
  - 95.23 + 5.56*(22) + 63.39*(1) + 197.43*(0) + 121.01*(0) = 281
- Forecast for period 22 (Spring of year 6) using the unadjusted model:
  - 171.07 + 7.43*(22) = 335
- substantial differences, even though the seasonal effect for spring is not as pronounced relative to winter (0) and summer (197.43)

Some words of warning

- Realised trend might be unsustainable
  - Sometimes a strong positive trend is expected, eg the introduction of a new successful product
  - but strong positive trends usually tail-off after a while
    - market saturation, product reaches maturity, competitive pressure, etc
- Time series predictions are always ‘out-of-sample’
  - by definition: future time periods have not yet been observed!
  - Therefore, the error of the prediction increases for more longer-term predictions.

- Due to these reasons, simple time series models as presented in this module are not suited for longer-term predictions
  - More advanced models that try to tackle these issues exist, but are more complicated and out of scope for this module