QBUS6840 Lecture 03

Decomposition Methods

Professor Junbin Gao

The University of Sydney Business School

Outlines

- Time series decomposition: Additive and multiplicative models
- Forecasting using decompositions
- X11 decomposition

Readings: Online text Chapter 6: otexts.com/fpp2/decomposition.html and/or BOK Section 6.3 (from pg 295), BOK Ch. 7

Time Series Decomposition

- Interpreting a time series

$$y_t = \text{patterns} + \text{error}$$

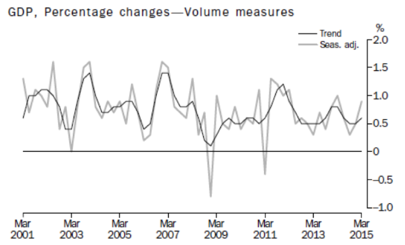
- Decomposition model

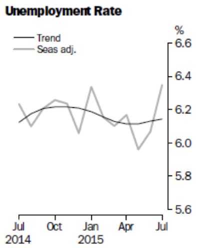
$$y_t = f(T_t, S_t, C_t, \epsilon_t)$$

- Pioneered by the French government in 1911.
- US Bureau of Census and Bureau of Labor Statistics.
- Formalised by Macauley (1931), The Smoothing of Time Series, New York: National Bureau of Economic Research.
- The goal is usually seasonal adjustment rather than forecasting.

Seasonal adjustment

- The Australian Bureau of Statistics (ABS) adjusts series such as: Building approvals, unemployment rate, labour force, change in gross domestic product, average weekly earnings, population growth (by state).



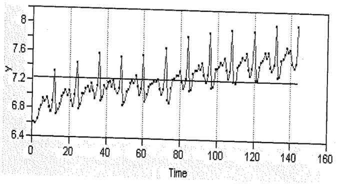


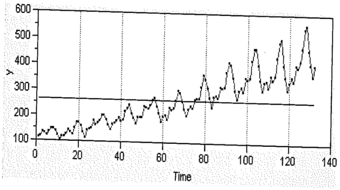
- Additive:

$$y_t = T_t + S_t + C_t + \epsilon_t$$

- Multiplicative

$$y_t = T_t \times S_t \times C_t \times \epsilon_t$$





Additive Model

- We can convert the multiplicative model into an additive model by noting that

$$y_t = T_t \times S_t \times C_t \times \epsilon_t$$

implies

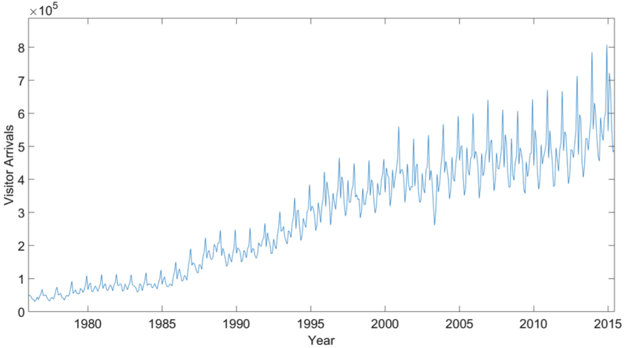
$$\log y_t = \log T_t + \log S_t + \log C_t + \log \epsilon_t$$

Some Notes

- Usually intuitive methods are applied to estimate these models
- Most often used when components or parameters are NOT changing over time
- Additive or multiplicative depends on the type of seasonal variation
- Additive if seasonal variation does not change much; multiplicative if the seasonal variation is proportional to the trend

Additive or Multiplicative?

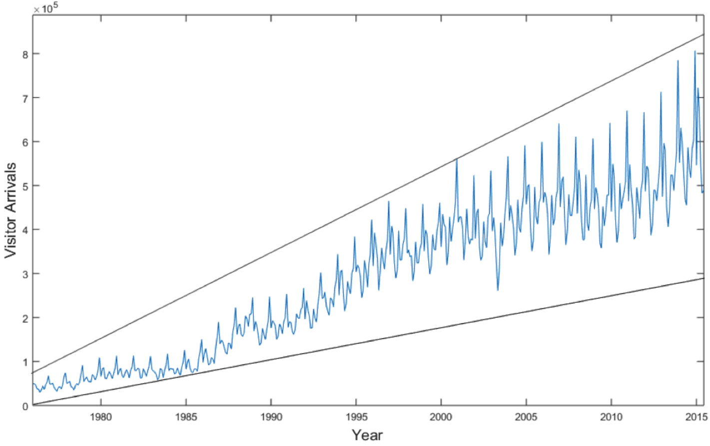
- Is the seasonal variation proportional to the trend?



Data from the Australian Bureau of Statistics

Additive or Multiplicative?

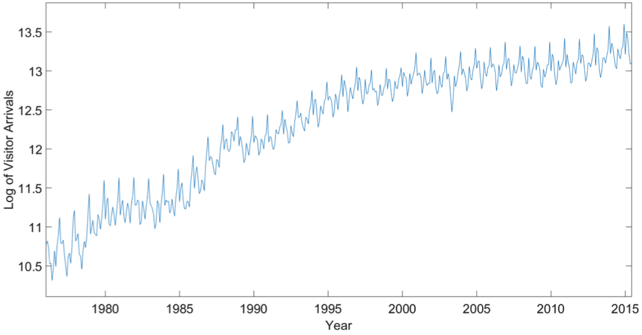
- A multiplicative model seems best for the visitor arrival series



- If this proportional is (linearly) shrinking, then what?

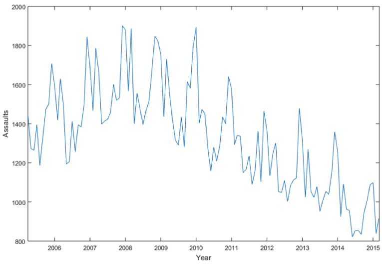
Additive or Multiplicative?

- Another way to visualize whether a multiplicative model is adequate is to plot the log series.



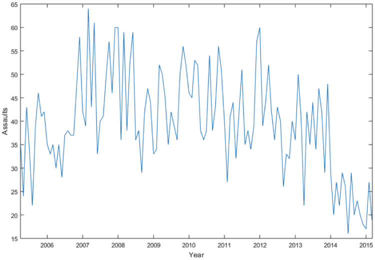
Again: Additive or multiplicative?

- Alcohol related assaults in NSW
- Data from the NSW Bureau of Crime Statistics and Research: www.bocsar.nsw.gov.au/Pages/bocsar_pages/Alcohol_Related_Violence.aspx



Again: Additive or multiplicative?

- Alcohol related assaults in King's Cross
- Data from the NSW Bureau of Crime Statistics and Research: www.bocsar.nsw.gov.au/Pages/bocsar_pages/Alcohol_Related_Violence.aspx



Cycle: Hard to Model

- Trend and cycle components are often combined:

  $$\text{Trend-cycle components } (TC_t)$$

- Or equivalently assume
  - $C_t = 0$ in Additive Model
  - $C_t = 1$ in Multiplicative Model

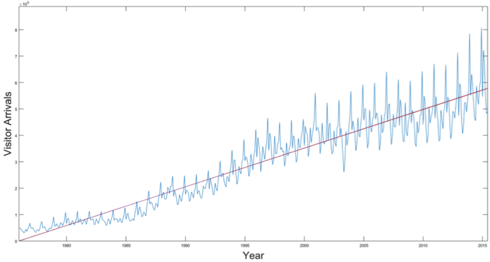
Trend Models

- Forecasting requires a parameter trend model. Common models are

$$T_t = \beta_0 + \beta_1 t$$
$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

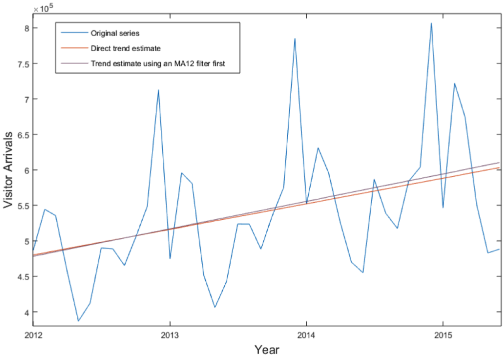
Estimating the Trend

- Suppose we want to estimate the trend using the linear model $T_t = \beta_0 + \beta_1 t$. We could simply regress Y_t by using T_t
- Regressing over Y_t may not be the best practice in some cases

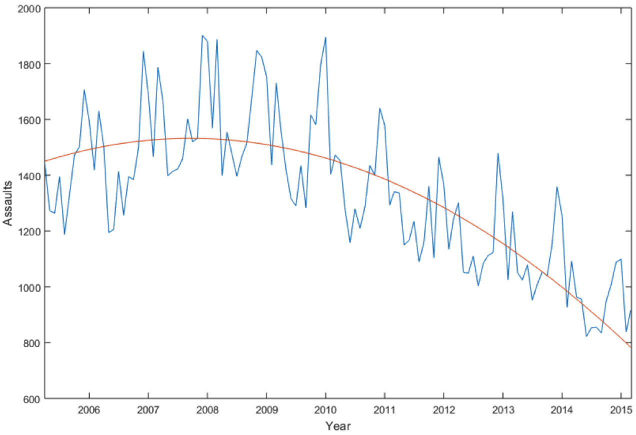


Estimating the Trend

- The accepted practice is to smooth the data to remove seasonal variation, and then regress smoothed values on time



Example: Fitting a quadratic trend



Notes on Trend

- Smooth to remove seasonality first.
- This reveals trend-cycle component.
- Also removes possible end and start effects.
- This is the initial trend estimate to only help estimate seasonality.
- We then re-estimate the trend later.
- Be very careful when extrapolating trend models for forecasting.

Estimating the Seasonal Component (Multiplicative)

- Model Assumption: $Y_t = T_t \times S_t \times C_t \times \epsilon_t \Rightarrow S_t \times \epsilon_t = \frac{Y_t}{T_t \times C_t}$
- The smoothed data (using for example an centred MA12) gives us the estimate $\widehat{T_t \times C_t}$ (notation for one value, not the product of T_t and C_t)

$$\widehat{S_t \times \epsilon_t} = \frac{Y_t}{\widehat{T_t \times C_t}}$$

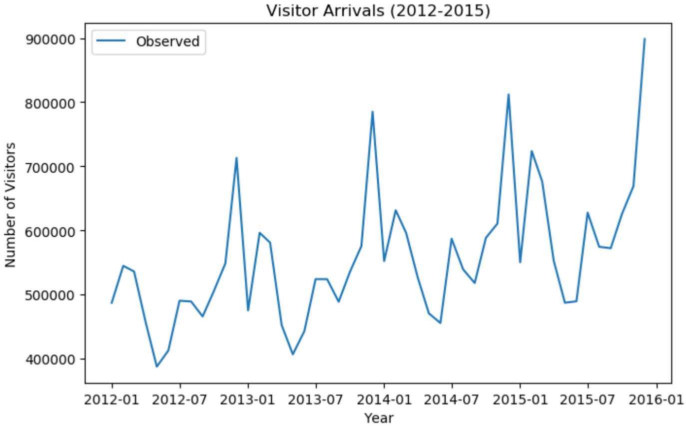
- The usual assumption (under monthly data) is that

$$S_t = S_{t-12} = S_{t+12} \Rightarrow \widehat{S}_t = \overline{S}_m$$

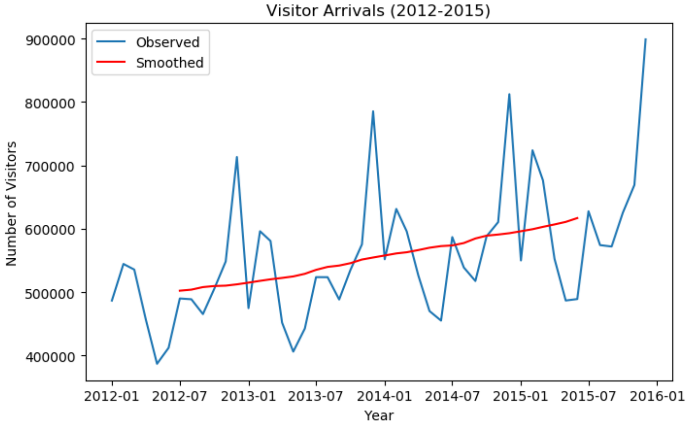
where $\overline{S}_m$ is called seasonal index and they ($m = 1, 2, ..., M$)are the normalized average of all observations in the month m of historical data, i.e.,

$$\sum_{m=1}^{M} \overline{S}_m = M.$$

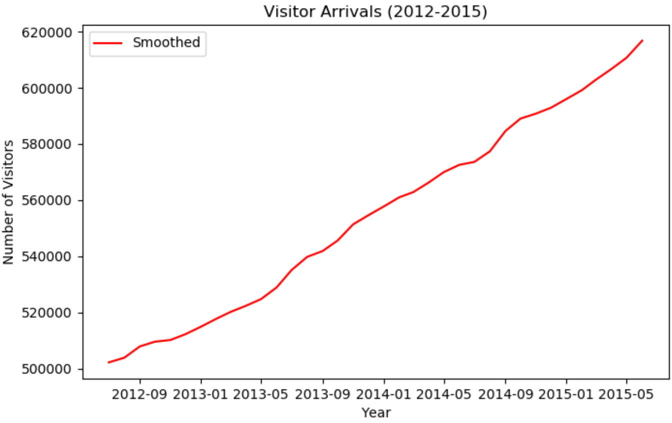
Example: The Time Series $\{Y_t\}$ (2012-2015)



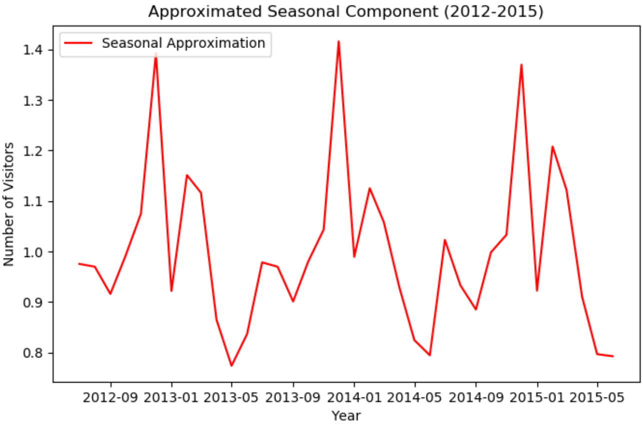
Example: Smoothing to have Trend-Cycle $\widehat{T_t \times C_t}$ (2012-2015)



Example: The (first-Approximated) Trend-Cycle $\widehat{T_t \times C_t}$ (2012-2015)



Example: Approximated Seasonal Series $\widehat{S_t \times \epsilon_t} = \frac{Y_t}{\widehat{T_t \times C_t}}$ (2012-2015)



How to convert this to an actual seasonal series?

Example: Calculation

- Use three years data to demonstration

Index	year1	year2	year3
Jul	0.975532	0.978562	1.02302
Aug	0.969948	0.969893	0.933196
Sep	0.916178	0.901172	0.885252
Oct	0.990678	0.980037	0.998444
Nov	1.07478	1.04341	1.03319
Dec	1.39218	1.41591	1.36988
Jan	0.921949	0.989406	0.922632
Feb	1.15115	1.12518	1.20782
Mar	1.11622	1.05857	1.12171
Apr	0.86493	0.929584	0.910456
May	0.773929	0.824791	0.796933
Jun	0.836715	0.79469	0.79285

Example: Calculation

- take the average of the three of the same months

Index	year1	year2	year3	Average
Jul	0.975532	0.978562	1.02302	0.992371
Aug	0.969948	0.969893	0.933196	0.957679
Sep	0.916178	0.901172	0.885252	0.900867
Oct	0.990678	0.980037	0.998444	0.98972
Nov	1.07478	1.04341	1.03319	1.05046
Dec	1.39218	1.41591	1.36988	1.39266
Jan	0.921949	0.989406	0.922632	0.944662
Feb	1.15115	1.12518	1.20782	1.16138
Mar	1.11622	1.05857	1.12171	1.09883
Apr	0.86493	0.929584	0.910456	0.901657
May	0.773929	0.824791	0.796933	0.798551
Jun	0.836715	0.79469	0.79285	0.808085

Example: Calculation

- Calculate the normalized constant by

$$c = \frac{12}{\text{sum of All month values}}$$
$$=12/(0.992 + \cdots + 0.808) = 1.00026$$

- Month indices
$\overline{S}_1 = 0.9446 * 1.00026 = 0.9449$, ...,
$\overline{S}_{12} = 1.3926 * *1.00026 = 1.3930$
- The seasonal component is

$$\{\widehat{S}_1, \widehat{S}_2, ..., \widehat{S}_{12}, \widehat{S}_{13}, ..., \widehat{S}_{24}, ...\}$$
$$=\{\overline{S}_1, \overline{S}_2, ..., \overline{S}_{12}, \overline{S}_1, \cdots, \overline{S}_{12}, \cdots\}$$

- For monthly data, there are only 12 different $\widehat{S}_t$ values. For example, when $t = 37$, we know this is a January, so $\widehat{S}_t = \overline{S}_{Jan}$ (ie. $m = 1$)

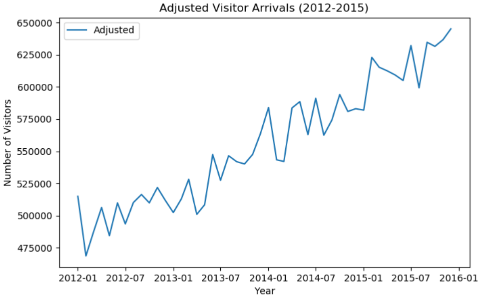
Index	year1	year2	year3	Average	Normalized
Jul	0.975532	0.978562	1.02302	0.992371	0.992626
Aug	0.969948	0.969893	0.933196	0.957679	0.957924
Sep	0.916178	0.901172	0.885252	0.900867	0.901098
Oct	0.990678	0.980037	0.998444	0.98972	0.989973
Nov	1.07478	1.04341	1.03319	1.05046	1.05073
Dec	1.39218	1.41591	1.36988	1.39266	1.39301
Jan	0.921949	0.989406	0.922632	0.944662	0.944904
Feb	1.15115	1.12518	1.20782	1.16138	1.16168
Mar	1.11622	1.05857	1.12171	1.09883	1.09912
Apr	0.86493	0.929584	0.910456	0.901657	0.901888
May	0.773929	0.824791	0.796933	0.798551	0.798756
Jun	0.836715	0.79469	0.79285	0.808085	0.808292

Seasonally Adjusted Series $T_t \widehat{\times C_t} \times \epsilon_t$ (one combined notation not three)

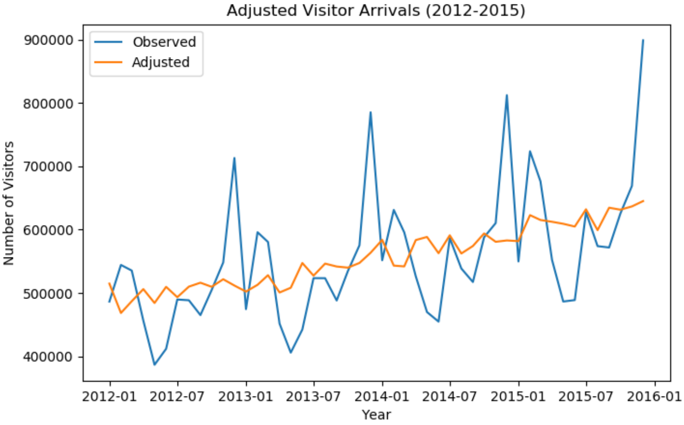
- We now calculate the seasonally adjusted series (taking off seasonal components)

$$Y_t = T_t \times S_t \times C_t \times \epsilon_t \Rightarrow$$

$$T_t \widehat{\times C_t} \times \epsilon_t = \frac{Y_t}{\widehat{S}_t}$$



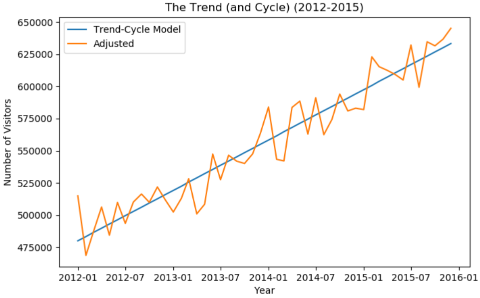
Adjusted Series against the Original Series



Re-estimating the Trend (-cycle)

- We regress seasonally adjusted data (in orange) on time to obtain the trend model
- Using which mathematical model depends on the shape of seasonally adjusted series. Here we have used a linear model

$$\begin{aligned} T_t \times \widehat{C_t} \times \epsilon_t &= \beta_0 + \beta_1 t + \epsilon_t \Rightarrow \\ \widehat{T_t} &= \widehat{\beta_0} + \widehat{\beta_1} t \\ &= 476759.25 + 3261.33t \end{aligned}$$



Forecasting Future Values (Multiplicative)

- We set $e_{t+h} = 1$ to make forecasts

$$\widehat{y}_{t+h} = \widehat{T}_{t+h} \times \widehat{C}_{t+h} \times \widehat{S}_{t+h}$$

if a well-defined cycle exists; otherwise, set $\widehat{C}_{t+h}$ to 1 to have

$$\widehat{y}_{t+h} = \widehat{T}_{t+h} \times \widehat{S}_{t+h}$$

if the cycle cannot be predicted.

Forecasting Future Values (Multiplicative)

- We set $e_{t+h} = 1$ to make forecasts

  $$\widehat{y}_{t+h} = \widehat{T}_{t+h} \times \widehat{C}_{t+h} \times \widehat{S}_{t+h}$$

  if a well-defined cycle exists; otherwise, set $\widehat{C}_{t+h}$ to 1 to have

  $$\widehat{y}_{t+h} = \widehat{T}_{t+h} \times \widehat{S}_{t+h}$$

  if the cycle cannot be predicted.

- For example, suppose $t = 48$-th month (Dec 2015) and we wish to forecast for Feb and May 2016, which means $h = 2, 5$. Hence $t + h = 50$ and 53 respectively. First use trend formula, e.g., calculate

  $$\widehat{T}_{t+h} = \widehat{T}_{48+2} = \widehat{\beta}_0 + \widehat{\beta}_1(48 + 2) = \beta_0 + 50 * \widehat{\beta}_1 = 639825.75,$$
  $$\widehat{T}_{t+h} = \widehat{T}_{48+5} = \widehat{\beta}_0 + 53 * \widehat{\beta}_1 = 649609.74$$

Forecasting Future Values (Multiplicative)

- As $t + h = 53$ means May 2016 and $t + h = 50$ Feb 2016, we will have $\widehat{S}_{50} = \overline{S}_{\text{Feb}} = \overline{S}_2$ and $\widehat{S}_{53} = \overline{S}_{\text{May}} = \overline{S}_5$ in the final forecast calculation
- The Final forecast for Feb 2016 is

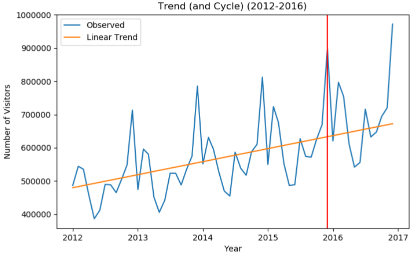
$$\widehat{Y}_{50} = \widehat{T}_{50} \times \widehat{S}_{50} = \widehat{T}_{50} \times \overline{S}_2 = 639825.75 * 1.162 = 743477$$

- The Final forecast for May 2016 is

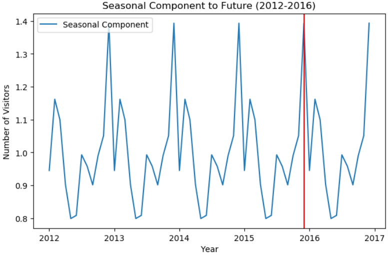
$$\widehat{Y}_{53} = \widehat{T}_{53} \times \widehat{S}_{53} = \widehat{T}_{53} \times \overline{S}_5 = 649609.74 * 0.798 = 518388$$

Forecasting T and S in 2016

Trend $\widehat{T}_{t+h}$



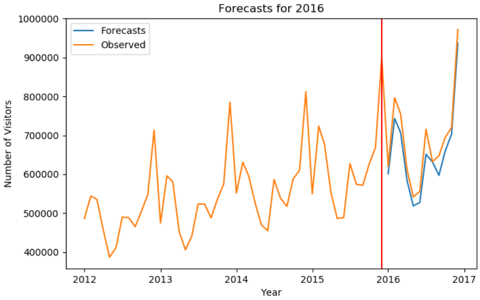
Seasonal $\widehat{S}_{t+h}$



Visitor Arrival Forecast

- We have used data from 2012 to 2015 to estimate the multiplicative model in which a linear model is used for the trend-cycle, and then forecast arrivals from Jan 2016 to Dec 2016 (in blue).

$$\widehat{T_t \times C_t} \times \epsilon_t = \beta_0 + \beta_1 t + \epsilon_t \Rightarrow \widehat{T_t} = \widehat{\beta}_0 + \widehat{\beta}_1 t$$

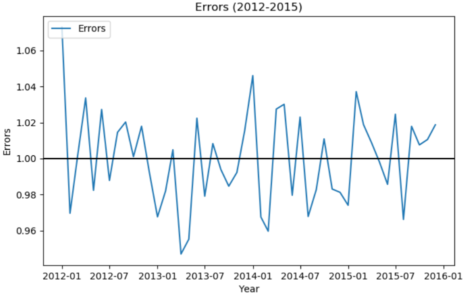


Does the model fit well?

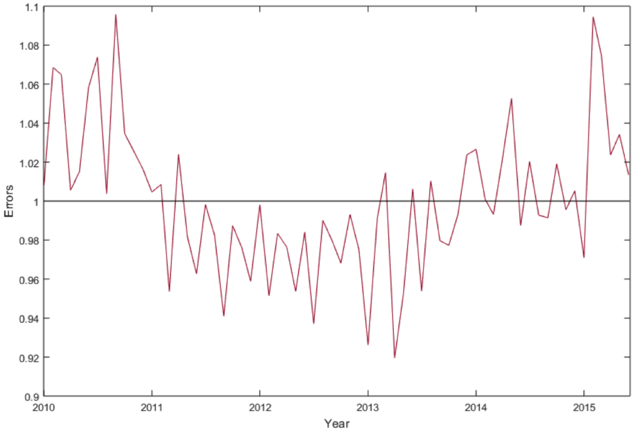
- Fit the multiplicative trend-seasonal model using the 2010-2015 data and compute

$$\hat{e}_t = \frac{Y_t}{\widehat{T}_t \times \widehat{S}_t}$$

- It seems okay. Ideally it should be noise around 1 (for additive model, this should be 0).



Is this model fitting well?



There are some patterns that have not been picked up by the trend and seasonal component. So this could be Cycle component!

Estimating the Cyclic Component (Multiplicative)

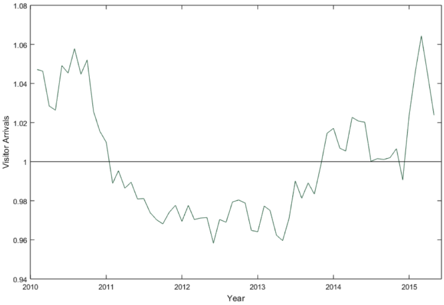
- We can estimate the cycle at each time t by first computing

$$\widehat{C_t \times \epsilon_t} = \frac{Y_t}{\widehat{T}_t \times \widehat{S}_t}$$

- And then smoothing, e.g., with an MA-3

$$\widehat{C}_t = \frac{1}{3}\left(\frac{Y_t}{\widehat{T}_t \times \widehat{S}_t} + \frac{Y_{t+1}}{\widehat{T}_{t+1} \times \widehat{S}_{t+1}} + \frac{Y_{t-1}}{\widehat{T}_{t-1} \times \widehat{S}_{t-1}}\right)$$

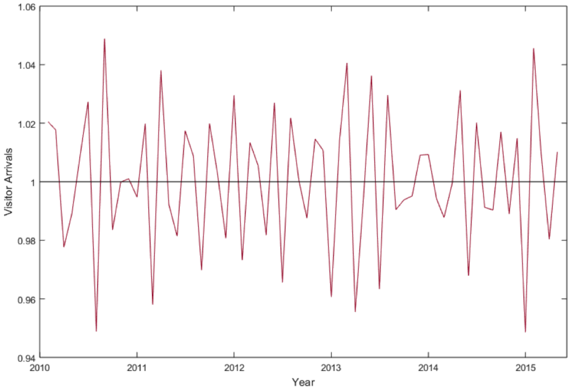
Estimating the Cyclic Component (Multiplicative)



- Why do we use an MA3 here? How about an MA12?
- If we are going to model this cycle, what model do you recommend?

Does the fit improve?

$$\widehat{\epsilon}_t = \frac{Y_t}{\widehat{T}_t \times \widehat{S}_t \times \widehat{C}_t}$$



Notes on the multiplicative decomposition (MD) model

- Lots of intuitive choices. No best method
- Different smoothers can be used at each step: WMA is often used
- Median smoothers can be used for robustness
- Outlier could be modelled and/or removed

"No Frills" MD Model

- Step 1: Smooth the data to remove seasonality, leading to the initial trend-cycle estimate. Denote this estimate by $\widehat{T_t \times C_t}$
- Step 2: De-trend the original series

$$\widehat{S_t \times \epsilon_t} = \frac{Y_t}{\widehat{T_t \times C_t}}$$

- Step 3: Estimate the seasonal indices ($\overline{S}_m$ Note: there are only M different values $\overline{S}_1, ..., \overline{S}_M$) based on $\widehat{S_t \times \epsilon_t}$ and normalisation, and then compute seasonally adjusted series:

$$T_t \widehat{\times C_t} \times \epsilon_t = \frac{Y_t}{\widehat{S}_t}$$

- Step 4: If forecasting, fit a trend model to $T_t \widehat{\times C_t} \times \epsilon_t$ to obtain $\widehat{T}_t$ otherwise smooth $T_t \widehat{\times C_t} \times \epsilon_t$ to obtain $\widehat{T}_t$

“No Frills” MD Model (cont.)

- Step 5: Estimate the cycle-error ($\widehat{C_t \times \epsilon_t}$) by

$$\widehat{C_t \times \epsilon_t} = \frac{Y_t}{\widehat{S}_t \times \widehat{T}_t}$$

- Step 6: Smooth $\widehat{C_t \times \epsilon_t}$ to estimate the cycle, obtaining $\widehat{C}_t$
- Step 7: Estimate the errors ($\widehat{\epsilon}_t$) by removing the cycle

$$\widehat{\epsilon}_t = \frac{Y_t}{\widehat{S}_t \times \widehat{T}_t \times \widehat{C}_t}$$

Algorithm for Additive Decomposition

- Step 1: Smooth the data to remove seasonality,e.g. a centred MA-12. This estimates the trend-cycle. Denote this estimate by $\widehat{T_t + C_t}$
- Step 2: De-trend the original series

$$\widehat{S_t + \epsilon_t} = Y_t - \widehat{T_t + C_t}$$

- Step 3: Estimate the seasonal indexes ($\widehat{S}_t$) based on $\widehat{S_t + \epsilon_t}$ and normalisation (i.e., making $\sum_{m=1}^{M} \overline{S}_m = 0$). Note: this can be done (1) find the mean μ of the initial seasonable indices. (2) subtract the mean from the initial seasonable indices to get the final $\overline{S}_m$ $(m = 1, 2, ..., M)$
- Step 4: Calculate the de-seasonalised observations

$$d_t = T_t \widehat{+ C_t +} \epsilon_t = Y_t - \widehat{S}_t$$

Algorithm for Additive Decomposition (cont.)

- Step 5: Smooth the de-seasonalised series to get $\widehat{T}_t$ for example by using the linear trend model

$$d_t = \beta_0 + \beta_1 t + \epsilon_t = T_t + \epsilon_t \Rightarrow \widehat{T}_t = \widehat{\beta}_0 + \widehat{\beta}_1 t$$

- Step 6: Estimate $\widehat{C}_t$ from a MA of $Y_t - \widehat{S}_t - \widehat{T}_t$
- Step 7: Estimate the errors ($\widehat{\epsilon}_t$) by removing the cycle

$$\widehat{\epsilon}_t = \widehat{C_t + \epsilon_t} - \widehat{C}_t = Y_t - \widehat{S}_t - \widehat{T}_t - \widehat{C}_t$$

The X11 Method

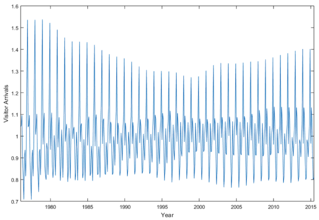
- X11 was developed by the U.S. Bureau of Census in the United States in 1965.
- Integrated into software packages such as SAS, R and EViews.
- It seasonally adjusts data and estimates the components of a time series.
- For the X11 method, there must be at least three years of observations in the input data sets.
- Unfortunately I have not found out a Python implementation of the X11 method.

X11 Algorithm Steps

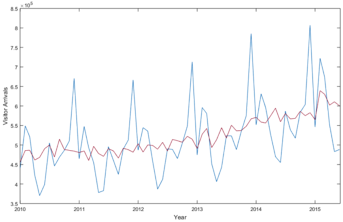
1. Centred MA12 to estimate trend-cycle, thus the series is de-trended to reveal seasonal factors.
2. Using the de-trended series, we apply an 3x3 MA filter to each month separately to smooth seasonal factors and give initial seasonal indices
3. The trend-cycle-error is estimated by removing seasonality using (2).
4. H9, 13 or 23 smoother (depending on the volatility) on trend-cycle-error estimates to obtain trend-cycle.
5. Repeat step (2) to get final seasonal indices.
6. Re-estimate trend-cycle-error as in (3) using the final St series, giving the adjusted data.
7. Final estimate of trend-cycle is H9 smoother after (6).
8. Final estimate of the irregular component

The detailed algorithm in Bleikh and Yong https://books.google.com.au/books?id=WfefCwAAQBAJ&pg=PP66&lpg=PP66&dq=The+X11+method&source=bl&ots=H_q3UTyeSH&sig=mv9clb6sIUqbNzzWk69X6pbBKxI&hl=en&sa=X&ved=0ahUKEwi4xP-gsrDLAhWIHpQKHRO3Chc4ChDoAQgaMAA#v=onepage&q=The%20X11%20method&f=false

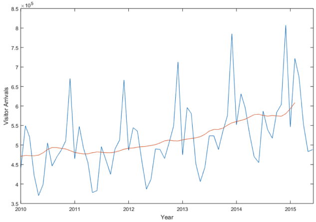
The X11 Method



X11 Seasonal Indexes $\widehat{S}_t$



X11 Seasonally Adjusted Series



X11 Trend Estimate

Summary

- It is fundamental to use genuine forecasts to evaluate the forecasting methods. Generating 'real time' forecasts is a good way to do this.
- You need to investigate whether an additive or multiplicative decomposition model is more appropriate for the data.
- Once you decide on the decomposition model, it is a matter of following the appropriate steps as we saw in this lecture. There is a lot of choice in the details.