Seasonal adjustment

"The first time is a blip, the second a coincidence, the third a trend", Anonymous

Standard plot: Visualising seasonality using R and ggplot2 (part 1)

Posted on **August 11, 2012**

When analysing data one of the best and most obvious things to do first is to plot it. This is simple and easy advice and you can plot data in many different ways. In recent years there has been what seems like an explosion in the philosophy of data visualisation (e.g. datavis, infographics, or any other fancy name). Suddenly anyone can turn themselves into a data visualisation wizard by taking some data, making some trendy little circles or boxes with curved edges, adding a few random colors, and then choosing a hand-written font and bingo! They (and everyone else) now think they are experts in finding patterns and analysing information.

To get on the band-wagon I thought it would be useful to illustrate one of the standard time series plots that can be used to help assess the properties of seasonality in a time series. These may not be as eye-candied as a typical data visualisation but at least they might serve a useful purpose.

To illustrate this I've download some data and also used R with the ggplot2 package. In doing this I found two frustrating things with using ggplot2. The first is that you need your data in a good format, which is typically an R dataset with all the variables you will need. In reality this can take the longest amount of time to sort out! The second is the logic of the whole ggplot2 thing and the terminology that has been used for the names and functions. It has taken me a lot of playing around to get to grips with it but once the basics are sorted it can produce nice looking graphics relatively easy.

For this illustration I've used the Australian Labour Force unemployment persons from the Australian Bureau of Statistics (ABS). You can get this data from here. One of the good things that the ABS do is that they publish the original, seasonally adjusted and the trend estimates. Having these components immediately available makes things easier as there is some small data derivation we need to do. In this example the irregular component has been derived by taking the seasonally adjusted data (which by definition is trend multiplied by the irregular component) divided by the published trend estimate to give the irregular component.

Using this dataset, and after some intense data manipulation, I finally got it into this form within R:

```
label     dat       period year
Seasonal  1.1241824      2 1978
Seasonal  1.0386693      3 1978
Seasonal  0.9935950      4 1978
Seasonal  0.9956933      5 1978
Seasonal  0.9683221      6 1978
Seasonal  0.9499924      7 1978
Seasonal  0.9494997      8 1978
Seasonal  0.9769546      9 1978
Seasonal  0.9321097     10 1978
Seasonal  0.9095337     11 1978
Seasonal  1.0728631     12 1978
…
Seasonal  1.0088955      5 2012
Seasonal  0.9573871      6 2012
Seasonal  0.9326108      7 2012
Irregular 1.0342590      2 1978
…
Irregular 1.0053110      7 2012
SI        1.1626958      2 1978
…
SI        0.9375638      7 2012
Mean      1.1168443      2 1978
…
Mean      0.9378250      7 2012
```

Note: The dataframe goes from February 1978 up to and including July 2012. The SI label reflects the seasonal multiplied by the irregular component.

Now that we have this, what can we use it for?

The best way to assess the seasonality in a monthly or quarterly economic time series is to use a seasonal and irregular chart. This groups the data according to periods, and plots the data within each period based on the yearly information. This will become clearer with the example below.

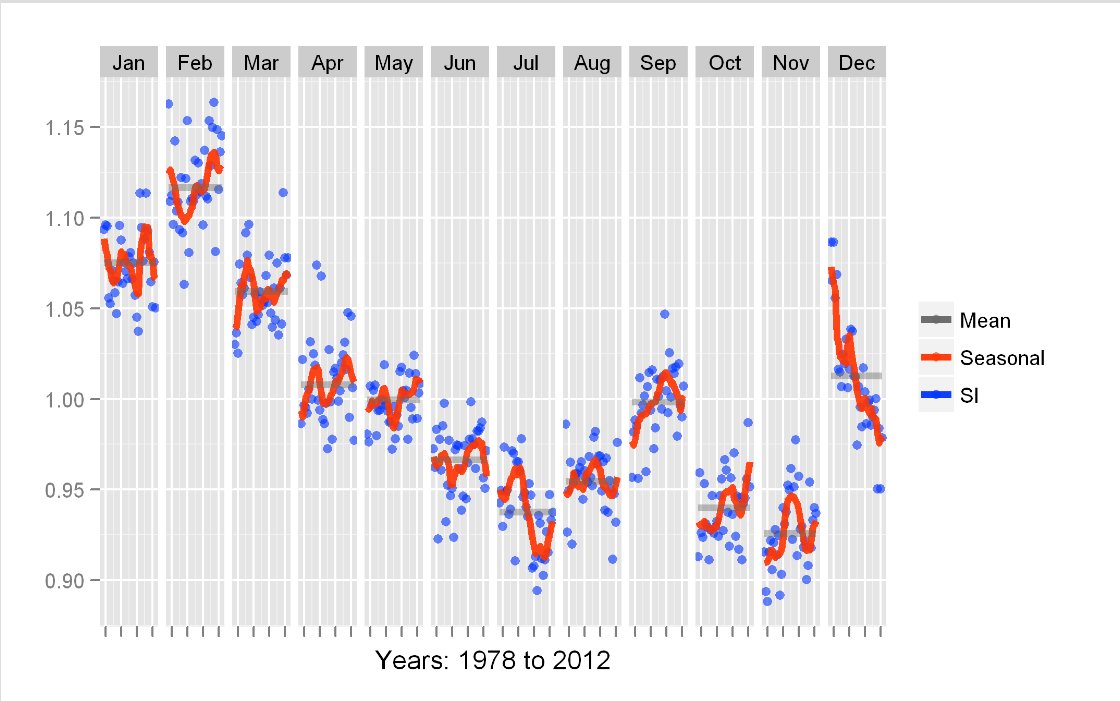
The benefit of this approach is that it highlights a little known or even acknowledged point by the hack analysts. Seasonality evolves over time and any estimation method for seasonal adjustment should capture this. A plot like the seasonal and irregular chart shows clearly how the estimates for the seasonality evolve over time. Note: this is why it is very very bad to use annual differences in the original estimates if you have any type of seasonality, particularly if it is evolving a lot.

Now for some R code. The following is probably not the most optimal code, but it does the job. Feel free to leave a comment if you can spot any obvious improvements! In this example the dataset is the data from above.

```
p1 <- ggplot(dataset,aes(year,dat,color=label)) +
geom_point(data=dataset[dataset[,"label"]=="SI",],size=I(2),alp
geom_line(data=dataset[dataset[,"label"]=="Seasonal",],size=I(1
geom_line(data=dataset[dataset[,"label"]=="Mean",],size=I(1.5),
opts(legend.title = theme_blank(),
axis.title.y=theme_blank(),
axis.text.x=theme_blank())

pout <- p1 + facet_grid(. ~ period) + xlab(paste("Years:",min(d
"to", max(dataset[,4])))
+ scale_color_manual(values=c("#666666", "#FF3300", "#0033FF" )
```

Typing “pout” in R then gives the following standard seasonal and irregular chart:

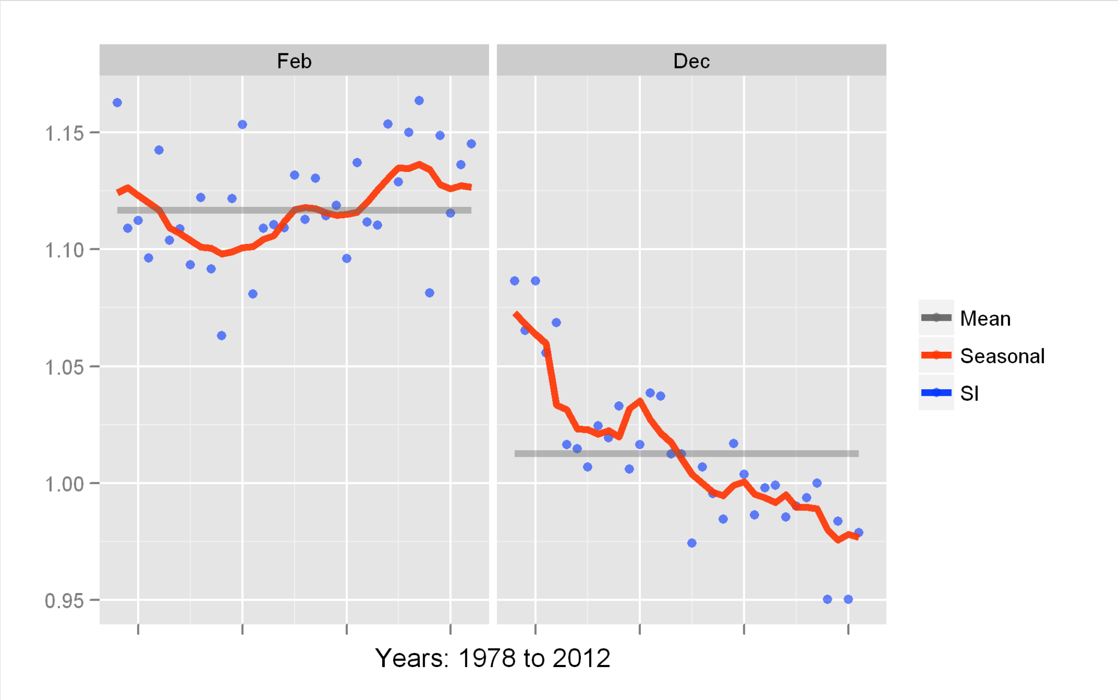


By grouping according to the periods (e.g. months or quarters), then each year can be displayed for each type of period. In this case, each blue dot represents a yearly estimate for the seasonal times irregular (blue), and the red line shows how the seasonal factor changes over time. The grey line gives the mean of the seasonal component over the whole time period.

To limit it to particular periods, it then becomes easy by just modifying the dataset. e.g.

```
dataset <- dataset[(dataset[,3]=="Feb")|(dataset[,3]=="Dec"),]
```

And re-running the code above to re-generate p1, and then pout gives:



For an analyst, this chart allows you to assess how well the seasonal factor is coping. What you could expect to see is that the seasonal factor (red line) should wander nicely through the seasonal and irregular time points (blue dots). If the seasonal factor is inadequate, then you could see blue dots show some sort of pattern, e.g. all above the seasonal factor, or a systematic occurence such as leap years or timing of Easter. If the seasonal and irregular time points are consistently above the seasonal factor, this indicates residual seasonality in the seasonally adjusted estimates which is a bad thing. These type of patterns are more appropriately tested using regression methods, but some of these type of impacts can show up visually if you know what you're looking for.

In this example, I'm not too worried about the actual data and giving an interpretation. In this case because some of the data has been indirectly derived, different estimates for the seasonal and irregular data could be derived if the time series was directly seasonally adjusted.

This example was more about showing how R and ggplot2 can be used to assess the outputs. In some coming posts I'll build on this example and show some different representations of the same outputs using ggplot2.

Share this:

This entry was posted in **R**, **Visualisation** and tagged **ggplot2**, **irregular**, **seasonal**, **unemployed** by **harrymac**. Bookmark the **permalink**.

[http://www.seasonaladjustment.com/2012/08/11/standard-plot-visualising-seasonality-using-r-and-ggplot2-part-1/] .

2 THOUGHTS ON "STANDARD PLOT: VISUALISING SEASONALITY USING R AND GGPLOT2 (PART 1)"

Pingback: Wheel plot: Visualising seasonality using R and ggplot2 (part 2) | Seasonal adjustment

Pingback: Clock plot: Visualising seasonality using R and ggplot2 (part 3) | Seasonal adjustment

Comments are closed.