B4 - Computer Numerical Analysis – Trade

B-CNA-410

Groundhog

Bootstrap

1.0.2

Groundhog

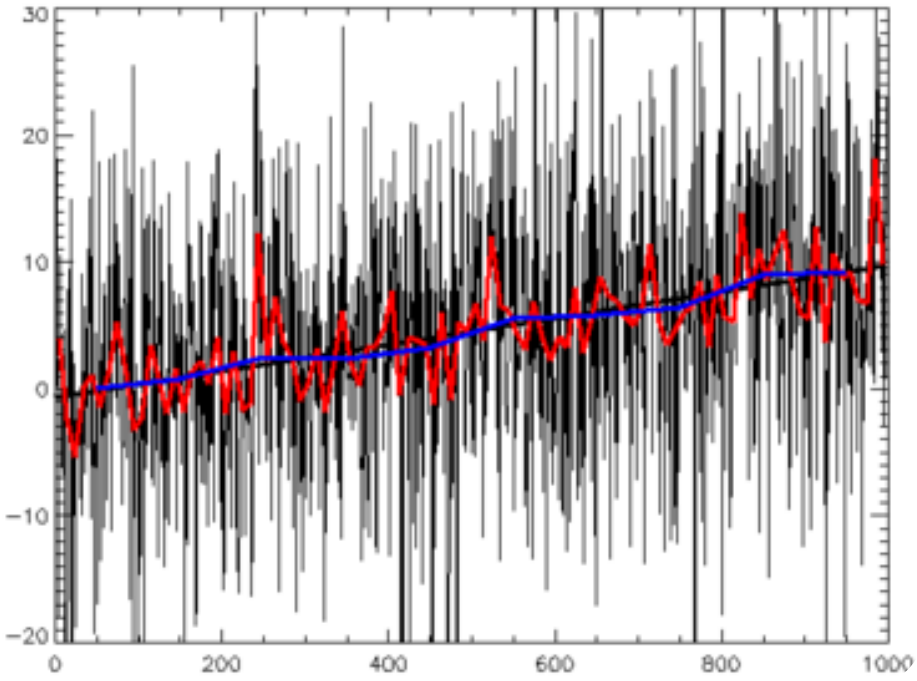
language: everything working on "the dump"
compilation: when necessary, via Makefile, including re, clean and fclean rules

- The totality of your source files, except all useless files (binary, temp files, obj files,...), must be included in your delivery.
- Error messages have to be written on the error output, and the program should then exit with the 84 error code (0 if there is no error).

In this bootstrap, you will learn how to deal with time series.
For now, you just need to understand the mathematical part.

No need to worry about error management.

First of all, you should try to understand the following figure:



Don't use functions/libraries that do the job for you.
That wouldn't make any sense.

STEP 0 - GETTING INDEXES

First of all, let's get indexes and store them in an appropriate container.

Terminal

```
~/B-CNA-410> cat -e indexes.txt
90.70
92.90
92.98
91.80
92.66
92.68
92.30
92.77
92.54
92.95
93.20
91.07
89.83
89.74
90.40
90.74
88.02
88.09
88.84
90.78
90.54
91.39
90.65
```

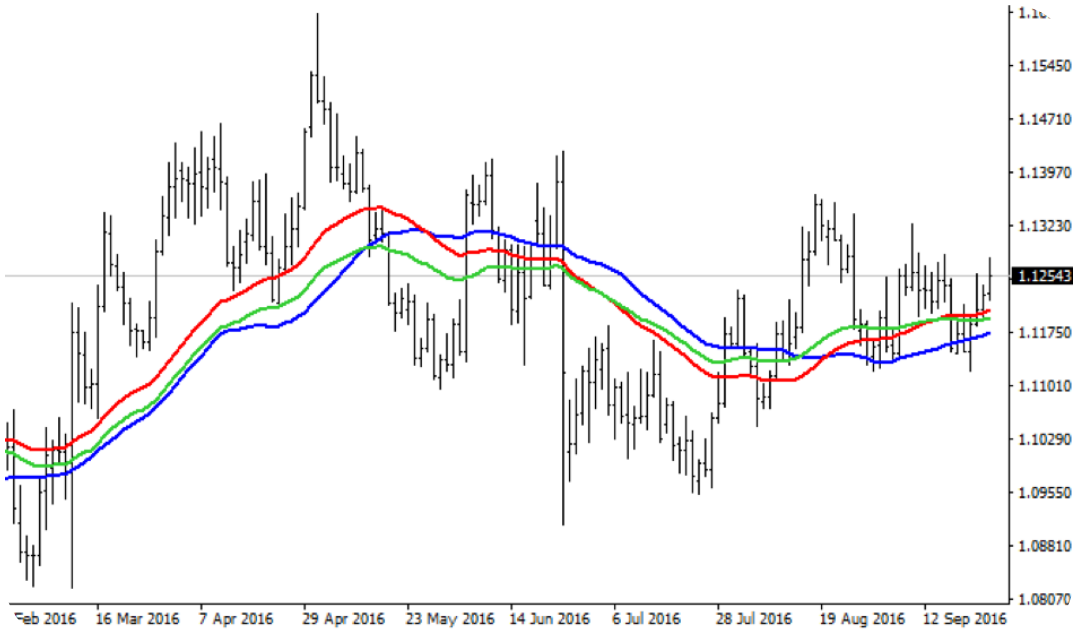
This history is important to predict the future.

Weather forecast is all about "trying to predict the future".

You probably want to visualize them as well. It might seem a big investment right now, but it will be profitable for sure.

Step 1 - Averages

The indexes you get are way too changing to be analysed as it is.
You need to pre-process the data by applying filters to it.
There are many interesting filters that can be applied with relevance to smooth the data.
Averages mainly.
Test some.
Test some more.
Test more some more.



Google [[[double] [exponential]] mobile] average

The **period** is the number of indexes used to compute an average (when computing an average).
Add it as a parameter of your program to compute the averages.
Make some tests.
Make some more tests.
Make more some more.

What is the impact of increasing/reducing the period?

You probably want to visualize the data. I know.

Step 2 - Standard Deviation

One of the most important mathematical measure is the standard deviation.
It is based on the difference between each index and the moving average and it assesses how far the indexes are from the mean (we talk about *dispersion*).

To be honest, the correct definition is:

> *The standard deviation is the square root of the sum of the squared difference between each index and the moving average on the period.*

If it is unclear for you, make your own research!

Compute the standard deviations and find a way to make these numbers readable.

You probably want to visualize the data and the curves you got. Once more.

Step 3 - Indicators

Now, you may want to produce more specific indicators that you would use for forecasting, and to detect a tendency switch (for instance)?
Is the latest value out of expected bounds?

Tendency switches

You can define a tendency as an average behavior over a period. This might be as simple as :

- the difference between first and last values
- the ratio between that difference and the average
- more basically the sign of that difference
- the number of days where temperature increased/decreased compared to previous days,…

A switch is a change of such a tendency, for instance when the tendency indicator:

- changes sign
- increases/decreases above a given threshold, or above a fraction of the standard deviation.

Test as many indicators as possible.
Test as many more indicators as possible.
Test more as many more indicators as possible.

Bands and Outliers

Data is assumed to vary around a given tendency.
The difference between actual signal and an ideal model is called noise, and cannot be totally removed.
Thus it is interesting to build bands around tendency, in between which all the regular values are expected to remain.
You can find various band estimation methods in the litterature.
Feel free to experiment.

God damn it, feel free!!

Outliers is a complementary technique.
Given a certain tendency, you want to spot data that does not fit it well, assuming it is artefact and should be discarded. Typical outliers will be outside of the bands you just defined.

What next?

Congratulations! You are now able to perform basic tendency analysis on time series.
To go a bit further, you should mix cleverly the most relevant indicators to build your own over-stunnig algorithm.