The Black Art of Smoothing

David I. Wilson
Department of Electrical & Electronic Engineering
Auckland University of Technology, New Zealand
email: diwilson@aut.ac.nz

Smooth your noisy process data with locally-weighted regression smoothers. They are reasonably fast, and simple to implement, and have only one tuning knob.

Cleaning up

Have you ever been embarrassed by the excessive noise in the plots displayed in your control room? Or perhaps you found it frustrating to extract meaningful trends from measured data? What you need is to clean up your data.

By 'cleaning up' I really mean smoothing, and such a topic lies in the murky underworld of data analysis. Consequently there is a marked reluctance for engineers to use this black art and for good reason. Such a cavalier approach is often considered dangerously close to cooking the data, there is a fear we may inadvertently suppress some of the key information in the data stream, or perhaps it is simply that we do not have any convenient smoothing algorithm at hand.

Conversely the reasons to smooth the data are that often we are interested only in the long term trends, or we might be smoothing prior to some post-processing such as peak height measurements in an online gas chromatograph, we may wish to remove spikes in the data, or we just may wish to use derivative action in our PID controller.

Once we have decided on what we are trying to uncover in our data, we need to select a smoother. Ideally we are looking for a fast, efficient, perhaps recursive implementation, we would like a single continuously adjustable smoothing parameter, and as an added bonus, we would welcome the ability to handle unequally spaced data and to extract derivatives.

To illustrate smoothing options, we need some time series data such as in Fig. 1. With the large collapse just visible on the extreme right, it looks like the recent share price for NZ Telecom, which, just coincidentally, in fact it is.

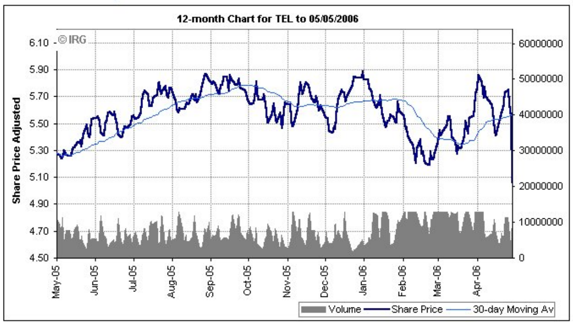

Figure 1. NZ Telecom's recent share price superimposed with a 30-day moving average filter. Data from Investment Research Group, IRG Ltd.

Smoothing locally

You probably already smooth your data using standard causal filters such as the 30 day moving average superimposed by the Investment Research Group in Fig. 1. However there smoothing alternatives *not* routinely used in the processing industries that we can borrow from econometrics or analytical chemistry. These are known collectively as locally weighted regression smoothing algorithms and historically received bad press due to the excessive computation required. But times have changed, and they deserve another look.

In this article I am going to demonstrate 3 families of smoothers: the Savitzky-Golay filter, the Hodrick-Prescott and Whittaker smoothers, and the loess smoother.

Of the three, the Savitzky-Golay (SG) filter is probably the most well known of the 'sliding window' type. Abraham Savitzky and Marcel Golay worked for the Perkin-Elmer Corporation and in the early 1960s were interested in efficient ways to both smooth the noise, but not inadvertently knock off the peaks in the spectroscopic data. The filters popularity is due to the fact that the algorithm is very slick for equally spaced data since it uses a sliding window concept much like a moving average filter, and can optionally deliver derivatives. The drawback is that the user can only adjust the window length, and interpolation order. Recommended values for the interpolating order is a modest 1 or 2, while the window length is in the same order of magnitude as for a moving average filter.

Compare this with the Hodrick-Prescott or Whittaker smoother family, [1]. These smoothers attempt to both fit a curve that faithfully represents the raw data, but is penalised if subsequent points vary too much. Mathematically this is a large, but sparse, optimisation problem which treats the entire data set at once, but the algorithm can be expressed succinctly in just three lines of MATLAB code. Both algorithms require the user to select a smoothing parameter, λ, that weights the two competing objectives. (Note the smoothed data is shifted in Fig. 2 to better illustrate the influence of the smoothing parameter, λ.) Unfortunately suitable values are really only found by trial and error, although 1600 is commonly recommended for 'quarterly financial data'.

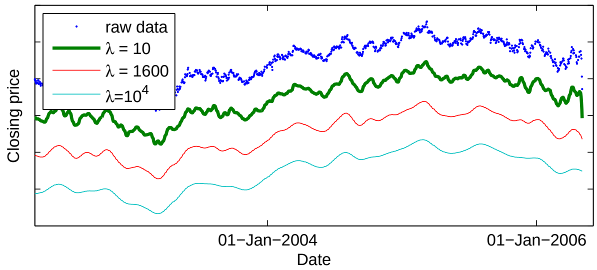

Figure 2. Hodrick-Prescott smoothers with different weights. (Note $\lambda = 1600$ is recommended.)

Finally we come to loess (or sometimes known as lowess, [2]) filtering illustrated in Fig. 3. Like the SG filter, this filter smooths only the data in a window which is slid through the data set. Like the Whittaker smoothers, we have a continuous tuning parameter, and unlike either, the tuning parameter α, is intuitive; it is the proportion of the full data set that we want to use in our window.

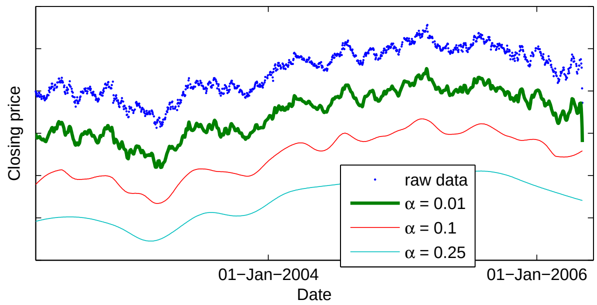

Figure 3. Loess smoother with different tuning factors α.

Which smoother is preferable depends on the application. If we want to capture the plummets, we will have to live with some noise. The SG filters suffer from end effects, but requires minimal storage. By considering all the data at once, the Whittaker smoothers avoid the end effect transients, but efficient application practically demands equally spaced data, and significant storage. For the speed/memory tradeoff, ease of tuning, and ability to handle unequally spaced data, the loess smoothers receive my vote for the method of choice. All the smoothing alternatives are compared in Fig. 4.

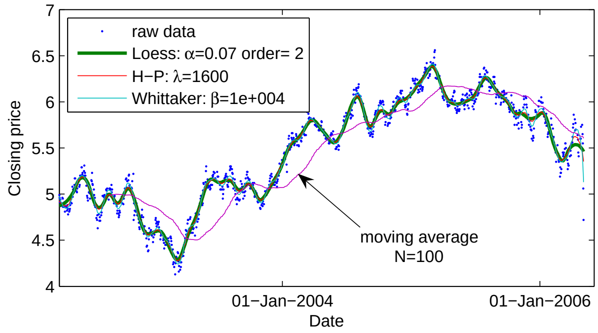

Figure 4. NZ Telecom's recent share price superimposed with various smoothers.

Trying it yourself

There is no substitute for experimenting with these smoothing algorithms on your own data. You will find the implementations of the smoothing routines discussed in this article and some sample data files at `http://homepages.ihug.co.nz/~deblight/smoothing.html`. There are MATLAB implementations, and for those without access to MATLAB, an impressive freeware equivalent, SCILAB, is downloadable from `www.scilab.org`.

Is it possible to smooth Telecom's share price to spot long term trends and still be alerted to possible plummets? Well, these are smoothers, not oracles, so don't cancel the morning paper just yet.

References

[1] Paul H.C. Eilers. A perfect smoother. *Analytical Chemistry*, 75(14):3631–3636, 2003.

[2] W. S. Cleveland. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician*, 35(54), 1981.