Following The Trend

Stephen Jewson, Risk Management Solutions*
and
Jeremy Penzer, London School of Economics

July 18, 2004

Abstract

We review recent research into the question of how to deal with the trends in meteorological data when pricing weather derivatives.

1 Introduction

Weather derivative pricing is often based on an analysis of historical meteorological data. The two most important choices when pricing from such data are a) how many years of data to use and b) how to deal with trends that may exist in the data. Given how much these choices affect the pricing of weather contracts it is surprising that there has been very little written on this subject. We now discuss the results of some of our recently published research which we believe sheds significant light on the pros and cons of the different choices one can make. We focus on the most straightforward question of how to estimate the fair level for the trading of uncapped swaps (the expected settlement index). Other questions such as how to estimate the whole distribution of future weather indices are also important, especially for the pricing of options: we hope to address them in a subsequent article.

2 Trend models

There are a number of different trend models in use, including using no trend at all (which we call 'flat line') removing a best fit linear trend (with or without extrapolation), and fitting more complex non-linear trends such as loess. A mixture of received wisdom, anecdotal evidence and intuition suggests that flat line can usefully be applied when only 10 years of data is being used, best fit linear trends can be applied to around 30 years of data and loess trends to around 50. But is there any way we can get a more scientifically based understanding of the properties of these different trends for different numbers of years of data? And could there be other models that would work better than any of these?

3 Backtesting of trends

One direct way to understand the performance of these different detrending methods is through a back-testing experiment which investigates how the different methods would have performed in the past. We describe such an experiment in [1]. A crucial aspect is how to judge the results: in [1] we consider the distribution of prediction errors from the different methods. In particular we look at three quantities: a) the mean prediction error, also known as the bias b) the standard deviation of the prediction error c) the root mean square prediction error

These measures need a little explanation, since understanding them is crucial to understanding the advantages and disadvantages of the different trend methods.

If the mean prediction error is zero then if we could repeat our prediction over and over again in independent tests the average error would tend towards zero. This is a nice property to have for any prediction system, but it is not sufficient on its own to guarantee that the predictions are any good. As an example of why it is not enough, predictions of daily temperatures in London that range from -100 to +100 could have a bias of zero. This example illustrates that we also need to consider the typical size of the errors.

* *Correspondence address*: RMS, 10 Eastcheap, London, EC3M 1AJ, UK. Email: `x@stephenjewson.com`

The standard deviation of the prediction error is one way to get information about the size of the errors. But again it is not sufficient on its own: the prediction that London temperatures are always -100 has a small standard deviation of errors but a huge bias, and would be a useless forecast. The standard deviation of prediction errors in a parametric trend model is made up of a combination of the variability of weather, the parameter uncertainty in the model and the model error (the extent to which the model is wrong). We note that the standard deviation of prediction errors is not necessarily the same as the standard deviation of the historical index values (detrended or not).

We have seen that neither mean nor standard deviation of prediction errors gives us the whole story. Ideally one would consider some measure that is a trade off between mean error and standard deviation of errors. There is no single best way to do this. We will consider the most obvious trade off, which is the root mean square error (RMSE). RMSE gives the typical size of prediction errors, and is related to the mean and standard deviation of errors very simply: it is the square root of the sum of the squares of the mean and standard deviation of the errors. Other trade offs would also be possible, and the particular trade off appropriate for use in weather derivative pricing might depend on the trading strategy being followed. For instance if one is selling many contracts in a non-competitive environment then reducing bias might be more important than reducing the standard deviation of errors.

Figure 1 shows the results from our backtesting study for winter HDDs, averaged over 200 US locations and the last 30 years, in terms of these 3 measures. In terms of mean error flat line is poor and gets worse the more data is used because it ignores the trend and hence overestimates the number of HDDs. Linear is much better and loess is near perfect. In terms of standard deviation of errors, however, flat line does much better. This is because it is a single parameter model. Linear trend does worse because it has two parameters, and the second parameter, for the slope of the line, is only very poorly estimated (the model is overfitted). Loess does even worse because it has even more parameters (implicitly) and is thus even more overfitted. Extrapolation of linear and loess trends makes these problems worse still. This is a salutary lesson in the dangers of overfitting. When we consider RMSE, we find that the 10 year flat line method does best: the bias may be high but this is compensated for by the low variance. 30 year linear does nearly as well: even though the variance is higher the bias is lower. If we extrapolate the lines by eye it would seem likely that linear would catch up with 10 year flat line at around 35 years. Loess does worse than flat line and linear, but again it is catching up and would maybe catch 10 year flat line at around 50 years if we could extend the study that far (unfortunately there is not enough historical data to do that).

4 The theory

Can we explain the results we see in this backtesting study? In two further papers ([2],[3]) we consider what results we would expect to see in the situation in which the real trend is linear. It turns out that the backtesting results are very close to this simple model. One of the most important lessons from these studies is that just because we think there is a trend does not mean that we should necessarily try and remove it using linear or loess detrending because of the severe danger of overfitting.

5 Local detail

There are a number of shortcomings in our backtesting study. One can argue that the future may not behave like the past, and certainly may not behave like the climate was behaving as much as 40 years ago. However, this assumption is difficult to avoid: the past is the only useful source of information that we have since climate prediction studies are as yet in their infancy when it comes to predicting local weather conditions. Another shortcoming is that we have averaged together results over different regions of the US and implicitly assumed that trends behave in the same way in the different regions. Is this true? Figure 2 shows the estimated linear trend slopes for 200 US locations using 30 years of data for winter HDDs, taken from [4]. What we see is that there are clear regional variations in the sizes of the trends. This has a number of implications. First, that one should repeat the backtesting study within each of these regions: one would not expect the same methods to work well in each region since the trends are so different. Secondly that there may be value in using trend information from surrounding stations: this may help reduce the problem of overfitting described above. Sorting out these two issues is an area of current research for us.

6 Making the right choice

Given all of this information about the performance of trends, how should we choose between different trend methods in practice? In [5] we discuss the possibility of using an automatic decision rule to choose the best detrending method for each situation. Initial results are promising, but more work is needed to test such rules on real data.

7 Optimal trends

A final question is whether there are other detrending methods that might perform better than flatline or linear, even when the real trend is linear. In fact there are, and with a bit of work one can derive the RMSE-optimal linear predictor of a linear trend. Perhaps surprisingly it is not the best fit linear trend, but is a blend of flat line and best fit linear trend. The blend depends on the strength of the trend, the size of the noise and the amount of data. The need for blending in this way is an example of a phenomenon that is well known in statistics as 'shrinkage'. It is also well known in the theory of meteorological forecasting and is often called 'tempering' or 'damping' of predictions. This optimal trend performs a trade-off between the low variance property of the flat line model and the low bias of the best fit linear trend model, and produces a prediction that beats both (see [6] for details).

8 Bibliography

[1] S Jewson and A Brix (2004). **Weather derivative pricing and the year ahead forecasting of temperature part 1: empirical results**; http://ssrn.com/abstract=535142.

[2] S Jewson (2004). **Weather derivative pricing and the year ahead forecasting of temperature part 2: theory**; http://ssrn.com/abstract=535143.

[3] S Jewson and J Penzer (2004). **Weather derivative pricing and closed-form expressions for the uncertainty from linear detrending**; http://ssrn.com/abstract=561841.

[4] S Jewson and A Brix (2004). **Weather derivative pricing and the spatial variability of US temperature trends**; http://ssrn.com/abstract=535924.

[5] S Jewson and J Penzer (2004). **Weather derivative pricing and a preliminary investigation into a decision rule for detrending**; in preparation.

[6] S Jewson and J Penzer (2004). **Optimal year ahead forecasting of temperature in the presence of a linear trend, and the pricing of weather derivatives**; http://ssrn.com/abstract=563943.

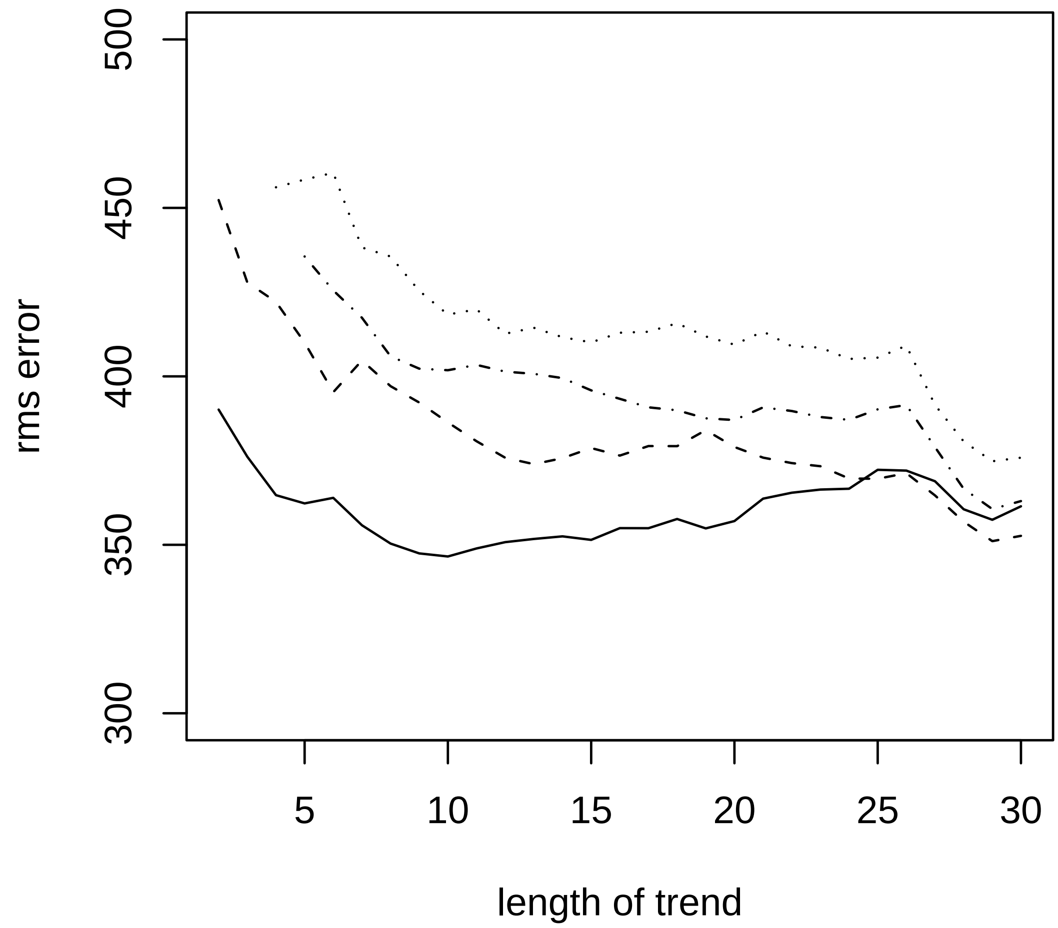

Figure 1: Results from the backtesting of various detrending methods, averaged over 50 years and 200 stations in the US. The methods used are flat line (solid line), linear (dashed line), loess 0.9 (dot-dashed line) and loess 0.6 (dotted line).

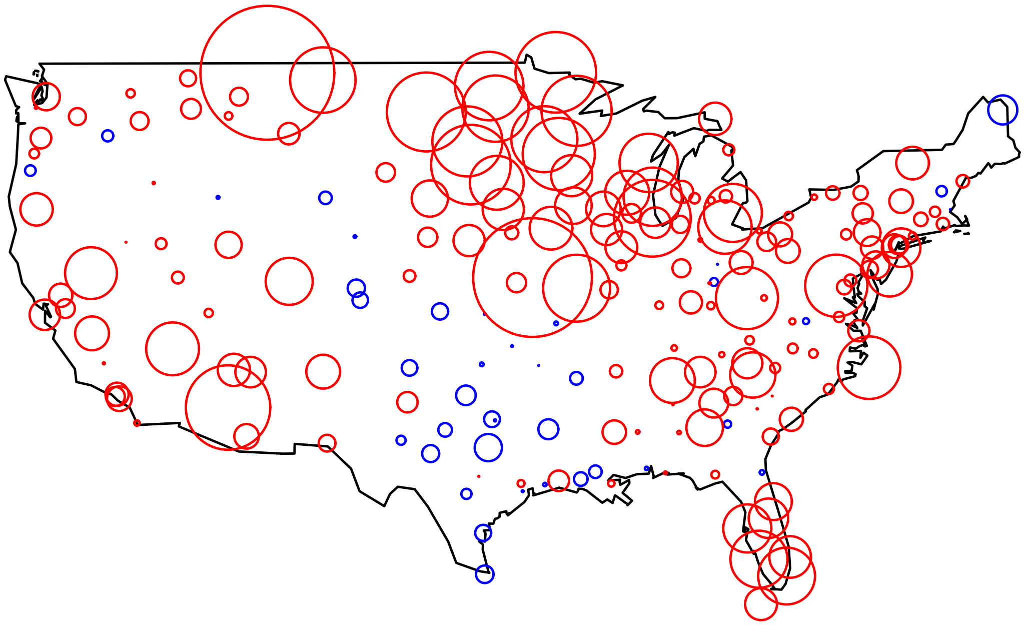

Figure 2: The sizes of observed temperature trends in winter average temperature in the US over the last 30 years. Red circles indicate positive trends, blue circles negative trends.