Is yesterday's model capable of explaining today's data?

I recall my father, back in the good old days, declaring confidently,

"Two minutes per tube stop",

to which he'd quietly add, "…more or less".

And there it was! At the impressionable age of 10, I encountered my first statistical model.

So what was Dad referring to? Well, when Dad, Mum and I would get on the London Underground, he'd estimate our journey time (y) by counting the number of stops (x) and multiplying by 2. His "more or less" disclaimer allowed for randomness (e). Were he pompous and statistically-minded (both of which his son has now become), he might have proclaimed instead,

"$y = 2x + e$".

Perhaps it's a good thing his son doesn't have children yet!

Of course, given that his model was just a caricature of reality (as all models are), it abstracted from various real-world complexities that could indeed result in shorter or longer journey times. As the celebrated statistician George Box once remarked, "all models are wrong, but some are useful". And Dad's model, albeit "wrong", was surprisingly useful. Twenty years later, I still use a variant of his formula to make back-of-the-envelope forecasts for my journey times.

Consider, however, that a lot can happen in twenty years: technological progress, economic shifts, policy changes, and a host of other occurrences. In other words, model parameters may change. Say, for example, that tomorrow the time taken per tube stop will rise permanently (for whatever reason) to 4 minutes. Clearly, Dad's model would no longer be relevant and I would be very late to work in the morning!

At risk of re-affirming my pomposity, I shall refer to tomorrow as a "changepoint". I shall also leave my poor Dad alone and move on to more serious examples of statistical models. For instance, we can consider sophisticated Dynamic Factor Models, used to extract key information from extremely large amounts of data.

Such models are used on a daily basis to influence monetary policy for the US, for Europe and for other countries around the world. Decisions based on these models affect mortgages, pensions, wages, exchange rates and price levels in general in a widespread manner. Such models are also used widely in the financial sector to make investments and allocate portfolios. Given the importance of advanced quantitative modelling techniques in some of the most vital aspects of decision-making in society, it is crucial that models used stay relevant and up-to-date.

It appears obvious then why the accurate and timely detection of changepoints (if any) in statistical models might be important. What if 9/11 happens tomorrow? What if people vote for Brexit? What if Mr.

Trump wins the election? Can society continue its reliance on yesterday's model to make decisions that affect our future?

My research, in particular, relates to the detection of changepoints in the afore-mentioned Dynamic Factor Models, and I hope I have sufficiently made the case for why research in this area might be necessary. In addition, if I may be so bold, I believe my research is also unique since it is one of the first to develop a solution to the question of how to detect changepoints in Dynamic Factor Models in real-time. That is, immediately when a changepoint is encountered (as opposed to a backward-looking point of view).

So, what's my secret? Well, I propose the use of a particular statistic. Let's call it Z.

For my statistically-minded readers, I note that Z is simply a ratio of eigenvalues crafted in a specific way. However, this detail is not so important. I only mention it to hint that computing Z is not very burdensome at all.

What is more important is that if we are using Dynamic Factor Models on a real-time basis (say daily) to forecast time series or for policy purposes, then I propose we also compute Z on a daily basis. Moreover, I propose that we monitor its value each day and watch for any irregularities.

The reason I make this suggestion is that I was able to provide mathematical proof that Z, my detection statistic, behaves in a very distinctive way when a changepoint is encountered. In particular, I proved that if today were a changepoint, the value of Z would spike upwards violently. However, if today were like any other day, Z would remain stable.

My finding motivates a real-time detection procedure. We declare a changepoint the first time Z breaches some pre-specified alarm threshold. (Of course, I also propose how to compute suitable alarm thresholds.)

To summarise, I believe that statistical models can and should be used to make sound decisions based on hard data. However, they must be used with care and, in particular, one must ensure that yesterday's model remains capable of explaining today's data.