Commentary on “Transparent modelling of influenza incidence”: Is the answer simplicity or adaptability?

Jennifer L. Castle[1]
Magdalen College and Climate Econometrics, University of Oxford

1. Introduction

It is a pleasure to comment on the paper by Katsikopoulos, Simsek, Buckmann and Gigerenzer (2020), where they present a provocative and stimulating viewpoint in which they argue that simple forecasting rules based on heuristics frequently outperform big data models and should be used as a benchmark when testing big data models. While testing more complex models against simple rules is uncontroversial and should be the norm in forecasting applications, I am more skeptical about the claim that simplicity delivers more accurate forecast performance. In my view the authors conflate simplicity with adaptability, consigning complex models to the forecasting bin as they are not ‘simple’. Instead, I argue that complex models could be highly adaptable to shifts in the data and may yield forecast gains over naïve devices. I elucidate below by describing why adaptable forecasting models are essential, reframing the debate from that of complex versus simple models to models that are adaptive to structural change versus those that have embedded equilibria. I examine the authors recency heuristic as a random walk model. Finally, I refute the notion that overfitting is a problem: overfitting can be controlled if implemented appropriately, before concluding in agreement with the authors that testing alternative models is essential.

2. Framing the complex/simple model debate in terms of adaptive models

Forecasting is easy in a stationary world. The conditional expectation will deliver the minimum mean-square forecast error so the best fitting in-sample model will produce the most accurate forecasts. However, when there are shifts in any aspect of the distribution, forecast theory derived for a stationary setting breaks down – anything goes.

As Hendry and Clements (2003) note, an unfortunate confusion which has resulted from the findings of past forecasting competitions is that ‘simpler models do better’.

Clements and Hendry (2001) explain that it is not simplicity that matters but adaptability to shifts in intercepts and trends. We do not have an agreed mathematical working definition of what is meant by complex, simple and big data models. If simple means fewer estimated parameters then one can conjure up many examples of cases where simple models forecast poorly. One of the simplest models imaginable is $y_t = \mu + \epsilon_t$, which is generally useless for data in levels, so a hidden assumption is the transformation being forecast. Differenced simple models may do well for growth rates but poorly for levels. The recent winner of the M4 forecasting competition, Smyl (2020), demonstrates that highly complex models utilizing big data can outperform. Big data techniques can help identify patterns in the data that were previously difficult to isolate, see Hassani and Silva (2015). The challenge to big data models is to avoid simple extrapolation of these patterns when they are subject to change. Combining new and innovative techniques for handling big data with methods to robustify forecasting models to structural change seems to be a fruitful avenue of research. Adapting big data forecasting models to wide-sense non-stationary data, such as the smoothed robust method proposed by Martinez et al. (2020), should address the criticisms of the authors.[2]

The Google Flu Trends forecasting model is an example of a lack of adaptability to non-stationary data. The authors show clear evidence that the recency heuristic dominates the Google Trends model, despite various iterations of the Google Trends model over time. This does not mean that the recency heuristic is the best forecasting model. Doornik (2009) finds that seasonality and calendar effects matter greatly for influenza forecasting. If the levels of influenza are higher than the norm at Thanksgiving, when levels are very low, then it is highly likely that there will be a much higher incidence over the forthcoming winter. It is the levels of influenza in the 'low period' (summer) that contains the most signal for the 'high period'. Furthermore, Doornik (2009) finds that search activity for 'cough', 'high fever' and 'child care'/'homework' has a positive impact on the percentage of visits for influenza-like illness, while 'school holidays' and 'tiredness' have a negative impact, so search data can indeed be useful.

The authors also argue that SIR models have failed in the Covid-19 pandemic due to lack of data availability on the parameters of interest, unreasonable model assumptions, and an inability to uniquely identify parameters, whereas simple heuristics do not face these challenges. The argument is more nuanced: Doornik et al. (2020) provide evidence that statistical models do outperform epidemiological models over certain periods of the pandemic, but I would caution against the use of simple heuristics too. The recency heuristic cannot hope to forecast during periods in which the epidemic is exponentially growing.

[2] I use the term 'wide-sense non-stationarity' to emphasize that every aspect of the distribution can shift, and avoid the common interpretation of non-stationarity to imply stochastic unit root behaviour.

There is strong evidence of seasonality (such as weekend effects) that also needs handling. Not only does the relative forecast performance of epidemiological models and statistical extrapolative models depend upon the point in the pandemic cycle, but simple heuristics cannot be used to assess various policy implications or undertake scenario analysis.

3. The Recency Heuristic as the Random Walk model

The most pernicious form of non-stationarity for forecasting is abrupt mean shifts, and this is when the random walk model comes into its own.

The proposed recency heuristic states that the prediction at time t, p_t, is given by the observed value of the CDC report in the previous week, o_{t-1}, so the prediction error is $e_t = o_t - p_t = o_t - o_{t-1} = \Delta o_t$, i.e. the random walk forecast.[3] The random walk is a simple model but is sophisticated in its implications.

Imagine a data generating process (DGP) y_t, which could be as complex or as simple as you like, but has an unconditional mean, $\mathsf{E}[y_t] = \mu$, that is constant over the in-sample period. At the forecast origin this unconditional mean shifts to $\mathsf{E}[y_{T+1}] = \mu^*$, but the shift is unpredictable. Now assume that the forecaster does not know the DGP specification or that the shift has occurred, so uses a misspecified forecasting model that aims to approximate the unknown DGP. No matter how well specified the forecasting model is in-sample, the forecaster will make a forecast error which will be dominated by the change in the unconditional mean $(\mu^* - \mu)$, and all other 'mistakes' in the forecasting model are likely to be small in comparison. There is systematic forecast bias which persists into the future, unless the shift is discovered and the forecasting model updated.

Now contrast that forecasting model with the random walk model, which states that the forecast made today will be what is observed yesterday. The forecaster will make a forecast error at the point of the mean shift, of magnitude $(\mu^* - \mu)$ in expectation, but the next period the random walk forecast will be based on μ^*, thus avoiding systematic forecast bias.

There is a more fundamental reason why the random walk forecast avoids systematic forecast failure. y_t captures every effect that the forecaster would like to know as it is the DGP. It contains all parameter changes, there are no omitted variables (even if some of the conditioning variables in the DGP are not measured) and no estimation is required. Using the previous realisation of y_t as the forecast has two drawbacks, namely the unwanted presence of a lagged error term which doubles the error variance, and the explanatory

[3] The authors define the prediction error as $e_t = p_t - o_t$, but I use the standard convention of outturn minus prediction, with the only impact being a sign change determining the interpretation of the forecast errors.

variables embodied in y_t are all lagged one period which introduces noise, but the random walk adapts to mean shifts as soon as they occur, making the recency heuristic sophisticated in its implications, see Clements and Hendry (1999) and Hendry (2006). This idea can be incorporated into big data models, see Martinez et al. (2020).

4. Overfitting should not be a problem – it can be controlled

There is a perceived concern that big data models overfit in-sample which leads to forecast failure. It is important to keep the two concepts distinct. Forecast failure can be a result of poor models or of good models that forecast badly because of unanticipated events that occur in the forecast period. Overfitting is described by Todd (1990) as fitting 'not only the most salient features of the historical data, which are often the stable, enduring relationships' but also 'features which often reflect merely accidental or random relationships that will not recur'. Overfitting is not a consequence of commencing with big data models, but an outcome of the selection strategy used. It is important to distinguish the number of parameters in the initial model (which could be huge) and the number of parameters in the final model (which could be very small). If very tight significance levels are used when applying selection to big data models there is no risk of overfitting. For example, if the initial set of regressors numbered 10,000 possible covariates for 100 observations of data, using a selection significance level of $\alpha = 0.0001$ would result in just one irrelevant variable being retained on average and 9,999 being discarded. Under a normal distribution the critical value is approximately 4, so the issue is retaining variables that are significant but have non-centralities less than 4, rather than the concern of overfitting, see Doornik and Hendry (2015). Big data models that overfit are not controlling the null retention frequency, but the charge that all big data models overfit and therefore forecast poorly is refuted.

5. Conclusion

In conclusion, the heuristics based on psychological theory of how people deal with rapidly changing situations that the authors highlight have parallels in econometrics. They are all adaptive to shifting data by removing any inherent mean embodied in the forecasting model. Assessing any forecasting model in terms of its adaptability to shifts and breaks will allow for both complex big data models and simple heuristics in the forecasting literature. The veracity of each will depend on the phenomena being forecast and the forecasting models used.

References

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Massachusetts: MIT Press.

Clements, M. P. and D. F. Hendry (2001). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting 17*, 550–554.

Doornik, J. A. (2009). Improving the timeliness of data on influenza-like illnesses using Google Trends. Mimeo, www.doornik.com/flu, Department of Economics, University of Oxford.

Doornik, J. A., J. L. Castle, and D. F. Hendry (2020). Short-term forecasting of the coronavirus pandemic. *International Journal of Forecasting*, In Press. DOI: 10.1016/j.ijforecast.2020.09.003.

Doornik, J. A. and D. F. Hendry (2015). Statistical model selection with big data. *Cogent Economics and Finance*, DOI:10.1080/23322039.2015.1045216. http://www.tandfonline.com/doi/full/10.1080/23322039.2015.1045216#.VYE5bUYsAsQ.

Hassani, H. and E. S. Silva (2015). Forecasting with big data: A review. *Annals of Data Science 2*, 5–19. https://doi.org/10.1007/s40745-015-0029-9.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics 135*, 399–426.

Hendry, D. F. and M. P. Clements (2003). Economic forecasting: Some lessons from recent research. *Economic Modelling 20*, 301–329. European Central Bank, Working Paper 82.

Martinez, A. B., J. L. Castle, and D. F. Hendry (2020). Smooth robust multi-horizon forecasts. Technical report, Department of Economics Working Paper, Oxford University.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting 36*(1), 75–85.

Todd, R. M. (1990). Improving economic forecasts with Bayesian vector autoregression. In C. W. J. Granger (Ed.), *Modelling Economic Series*, Chapter 10. Oxford: Clarendon Press.