High-dimensional, quantile forecasting with ensembles

Rob J Hyndman   23 July 2020

Quantile forecasting

- Almost everyone needs probabilistic forecasts whether they realise it or not.
- Tools for measuring the accuracy of probabilistic forecasts need to be more widely adopted and understood. CRPS and log scores as two examples.
- Tools for handling both probabilistic forecasts and high-dimensional grouped time series are just starting to be developed in the methodological literature. Early adopters have an advantage.

Quantile (or percentile) forecasting is not difficult. Any statistical forecasting method can be used to produce quantile forecasts by simulation. Suppose we are interested in forecasting the total sales in Australian cafes and we train an ETS model and an ARIMA model (Hyndman & Athanasopoulos, 2020) on the data to the end of 2018. Then we can simulate sample paths from these models to obtain many possible "futures". Figure 1 shows the last four years of training data and 3 futures generated from each of the two fitted models.
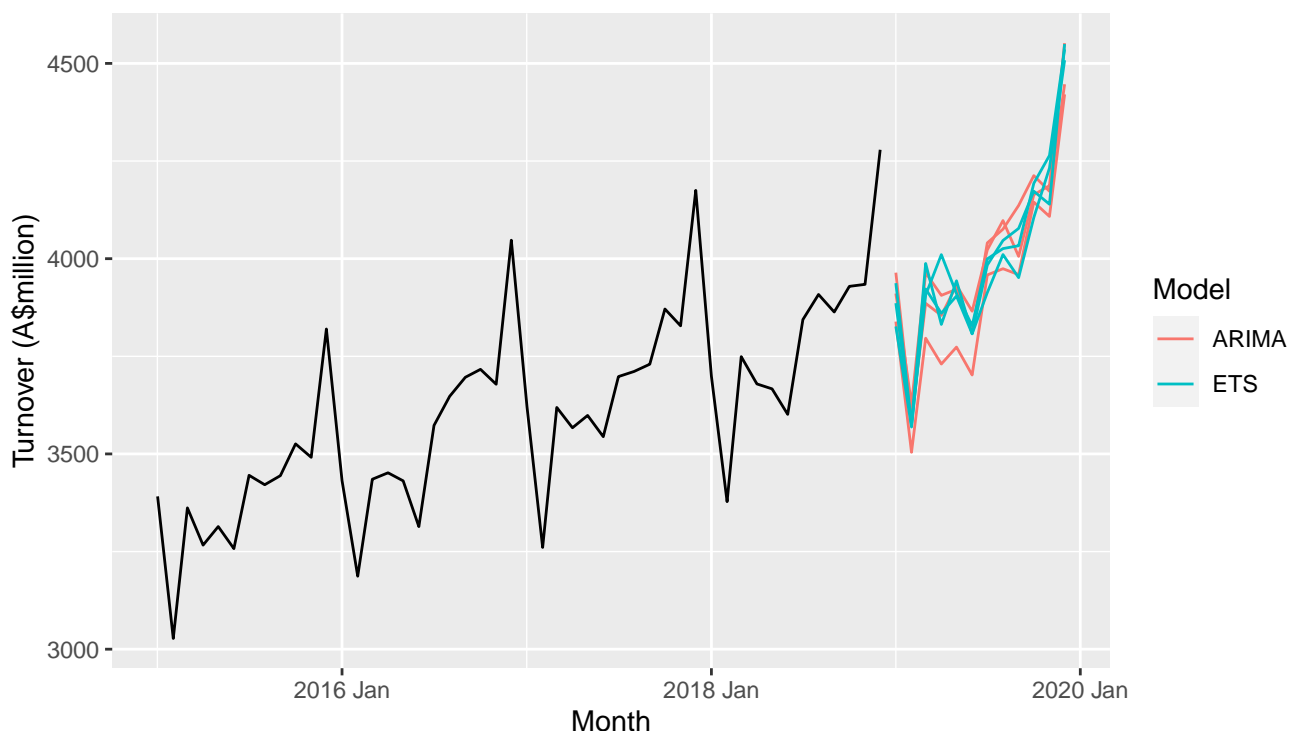
Figure 1: *Future sample paths obtained using an ARIMA model and an ETS model for the Australian monthly cafe turnover.*

If we repeat this procedure thousands of times for each model, we can obtain a very clear picture of the probability distribution for each future time period. The means of these distributions are the traditional point forecasts. Traditional 95% prediction intervals are equivalent to finding the middle 95% of the futures at each forecast horizon. Using simulations gives us the ability to easily compute any other quantiles. Figure 2 shows the deciles for the ETS forecasts (i.e., the 10th, 20th, ..., 90th percentiles).

Quantile forecasts also allow us to answer many more interesting questions. For example, we may wish to find prediction intervals for the total turnover for the next 12 months. This is surprisingly difficult to handle analytically but trivial using simulations — we just need to add up the turnover for each of the simulated sample paths, and then find the relevant quantiles. We might also want to find forecast the maximum turnover in any month over the next year. Again, that is a difficult problem analytically, but very easy using simulations. I expect that simulating future sample paths will play an
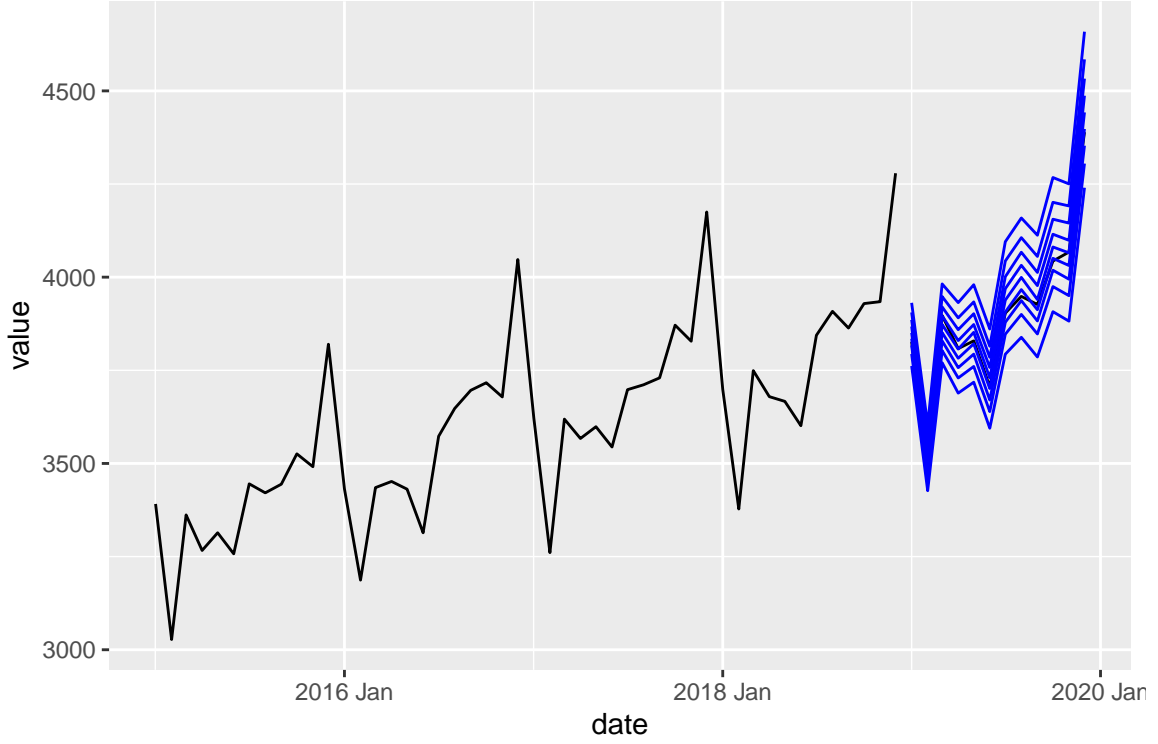
Figure 2: *Deciles for the ETS forecasts for the Australian monthly cafe turnover.*

increasingly important role in forecasting practice because it makes difficult problems relatively easy, and allows us to explore what the future might be like in ways that would otherwise be almost impossible.

Most business doing forecasting will be familiar with computing accuracy measures for point forecasts such as MAPE or RMSE values. With quantile forecasts, we need to use some alternative measures. Quantile scores provides a measure of accuracy for each quantile of interest. For example, suppose we are interested in the quantile with probability p for the future time periods, and let this be denoted by $q^p_{T+h|T}$, meaning the estimated quantile for time $T + h$ made at time T. That is, we expect the observation to be less than this value with probability p. For example, an estimate of the 95th percentile would be $q^{0.95}_{T+h|T}$. If y_{T+h} denotes the observation at time $T + h$, then the quantile score is

$$Q_p = \begin{cases} (1 - p)\left(q^p_{T+h|T} - y_{T+h}\right), & \text{if } y_{T+h} < q^p_{T+h|T} \\ p\left(y_{T+h} - q^p_{T+h|T}\right), & \text{if } y_{T+h} \geq q^p_{T+h|T} \end{cases}$$

This is sometimes called the "pinball loss function" because a graph of it resembles a pinball table. If $p > 0.5$, it gives a heavier penalty when the observation is greater than the estimated quantile than when the observation is less than the estimated quantile. The reverse is true for $p < 0.5$. A low value of Q_p indicates a better estimate of the quantile. Usually we would want to average the Q_p values of many forecast horizons in order to assess whether our model is estimating the quantiles accurately.

Often we are interested in the whole forecasting distribution, and then we can average the quantile scores over all values of p. This gives what is known as the "Continuous Ranked Probability Score" or CRPS. (For technical reasons, CRPS is usually reported as twice the average quantile score.)

In the Australian cafe example, we can compute the CRPS values over the 12 months of 2019 for each of the ARIMA and ETS models. To make it more interpretable, we can also compute the CRPS for a simple seasonal naive model, and then we can calculate the "skill score" equal to the percentage improvement for ARIMA and ETS over seasonal naive.

Model	CRPS	Skill score
ARIMA	39.1	43.1
ETS	31.5	54.0
SNAIVE	68.6	0.0

Here, ETS is providing the best quantile forecasts with a skill score of 54.0.

Ensemble forecasting

For more than 50 years we have known that ensemble forecasting improves forecast accuracy (Bates & Granger, 1969; Clemen, 1989). Also known as "combination forecasting", ensembles involve using multiple models and combining the results to produce the final forecast. There are several reasons why ensembles work.

- No model is perfect, and the data did not come from a model. As George Box has put it, "all models are wrong, but some are useful." Ensembles allow the good features of various models to be included, while reducing the impact of any specific model.
- Ensembles decrease the variance of the forecasts (Hibon & Evgeniou, 2005) by reducing the uncertainty associated with selecting a particular model.

Ensembling with probabilistic forecasting is easy via simulated future sample paths.

High-dimensional forecasting

- Many problems involve thousands of time series with some kind of grouping structure.
- We need software tools and models that are designed to handle these kinds of data.
- Most existing tools are set up for univariate time series, or multivariate time series. But the business world is almost always more complicated than that.
- Modern time series data bases use a time index and a key structure. – reference tsibble paper.

References

Bates, JM & CWJ Granger (1969). The Combination of Forecasts. *The Journal of the Operational Research Society* **20**(4), 451–468.

Clemen, R (1989). Combining forecasts: a review and annotated bibliography with discussion. *International journal of forecasting* **5**, 559–608.

Hibon, M & T Evgeniou (2005). To combine or not to combine: selecting among forecasts and their combinations. *International journal of forecasting* **21**(1), 15–24.

Hyndman, RJ & G Athanasopoulos (2020). *Forecasting: principles and practice*. 2nd edition. Melbourne, Australia: OTexts. OTexts.org/fpp3.