***Attribution for Forecasting: Principles and Practice***

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on February 7, 2020.

All forecasting models examined in FPP2 will concern prediction of data at future times using observations collected in the past. There will be no examination of cross-sectional prediction. FPP2 will explore the most reliable methods for producing forecasts with an emphasis on those methods that are replicable, testable, and have been shown to work.

***R packages used in FPP2***: (1) fpp2 v2.3; (2) forecast v8.7; (3) ggplot2

***Inspiration:*** Forecasting is a difficult activity and businesses that do it well have a big advantage over those whose forecasts fail.

Forecasting is required in many circumstances and time horizons and is an important aid to effective and efficient planning. Some things are easier to forecast than others and the predictability of events or quantities depends on several factors including:

1. how well we understand the factors that contribute to it;
2. how much data is available;
3. whether the forecasts can affect the thing we are trying to forecast.

The last factor underlies the efficient market hypothesis. When one or several of these factors are not satisfied, forecasters need to be aware of their own limitations, and not claim more than is possible.

***Concept:*** Oftentimes in forecasting, a key step is knowing when something can be forecast accurately, and when forecasts will be no better than tossing a coin. Good forecasts capture the genuine pattern and relationships which exist in the historical data, but do not replicate past events that will not occur again.

***Concept:*** Forecasters need to be able to differentiate between random fluctuations in the past data that should be ignored, and a genuine pattern that should be modelled and extrapolated.

It is wrong to assume that forecasts are not possible in a changing environment. In fact, good forecasting models capture the way in which things are changing and thus a changing environment is a fundamental assumption. Good forecasting models do assume that the way in which the environment is changing will continue into the future.

***Forecasting, Goals, and Planning***

***Forecasting:*** is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.

***Goals:*** are what you would like to happen. Goals should serve as a bridge between forecasting and planning. However, too often goals are set without any plan for how to achieve them and no forecasts for whether they are realistic.

***Planning:*** is a response to forecast and goals. Planning involves determining the appropriate actions that are required to ensure that you goals and forecasts are aligned.

***Things to consider when implementing forecasting systems***

***Identification:*** expertise in identifying what is and is not a good forecasting problem is important.

***Diversity:*** a forecasting system should a range of forecasting methods.

***Selection:*** choosing an appropriate method for each problem is a key to successful forecasting.

***Evaluation:*** always evaluating and refining forecasting methods is critical.

***Support:*** strong organizational support is paramount to successful implementation and utilization.

**There are several key considerations in the early stages of a forecasting project:**

1. deciding on what should be forecast and why it is being forecasted;
2. considering the forecasting horizon since the horizon affects forecasting model selection;
3. determining how frequently the forecast is needed affects the degree of automation;
4. incorporating the requirements and perspectives of the people that will use the forecast;
5. assessing the effort required to identify, gather, systematize, and clean the required data.

Selecting between the two primary forecasting paradigms (qualitative or judgmental versus quantitative forecasting) depends on whether the following two conditions are satisfied:

* numerical information about the past is available – concretely this is time series data that are observed at regular intervals of time;
* it is reasonable to assume that some aspects of the past pattern will continue into the future.

The unique properties, accuracies, and costs associated with each quantitative forecasting method influences method selection. Remember the ultimate goal – when forecasting time series data, the aim is to estimate how the sequence of observations will continue into the future with a given degree of confidence as specified by the prediction interval.

***Time Series Models:*** the simplest time series forecasting methods use only information on the variable to be forecast and make no attempt to discover the factors that affect its behavior. Time series models follow the general form:

***Explanatory Models:*** on the flip side, there are time series forecasting methods that do incorporate other factors or predictor variables.

The “error” term captures several things. There will always be changes in the target variable that cannot be accounted for by predictor variables. The error term allows for random variation and the effects of relevant variables that are not included in the model.

***Mixed Models:*** combine the methods of time series and explanatory models and include dynamic regression, panel data, longitudinal, transfer function, and linear system models.

There are various considerations that affect choosing between time series, explanatory, and mixed models. These considerations include:

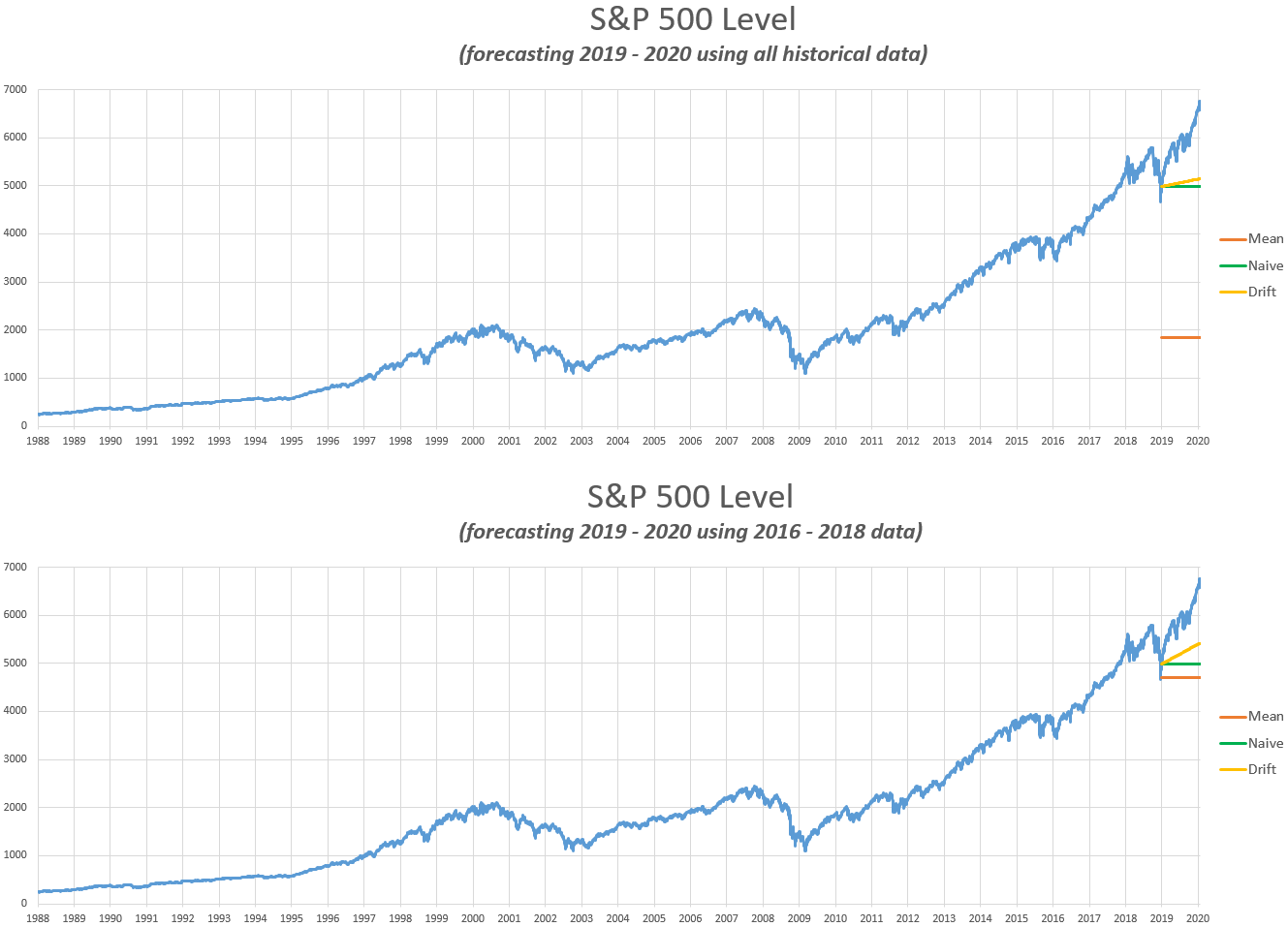
1. the overall system may not be well understood and even when it is, it may be extremely difficult to measure the relationships that are assumed to govern its behavior;
2. it may be difficult to forecast the future values of the various predictors as a precursor to being able to forecast the variable of interest;
3. the main concern may be only to predict what will happen and not worry about why it happens;
4. the forecast accuracy of time series models may be better explanatory or mixed models;
5. resources, data availability, and use cases are important practical issues.

**Chapter 3 – The Forecaster’s Toolbox**

§3.1 Some Simple Forecasting Methods

Some forecasting methods are very simple and surprisingly effective. These simple methods include:

1. Mean
2. Naïve
3. Seasonal Naïve
4. Drift

**Figure 1** The mean, naïve, and drift methods used to forecast the S&P 500 level for differing historical data timeframes.

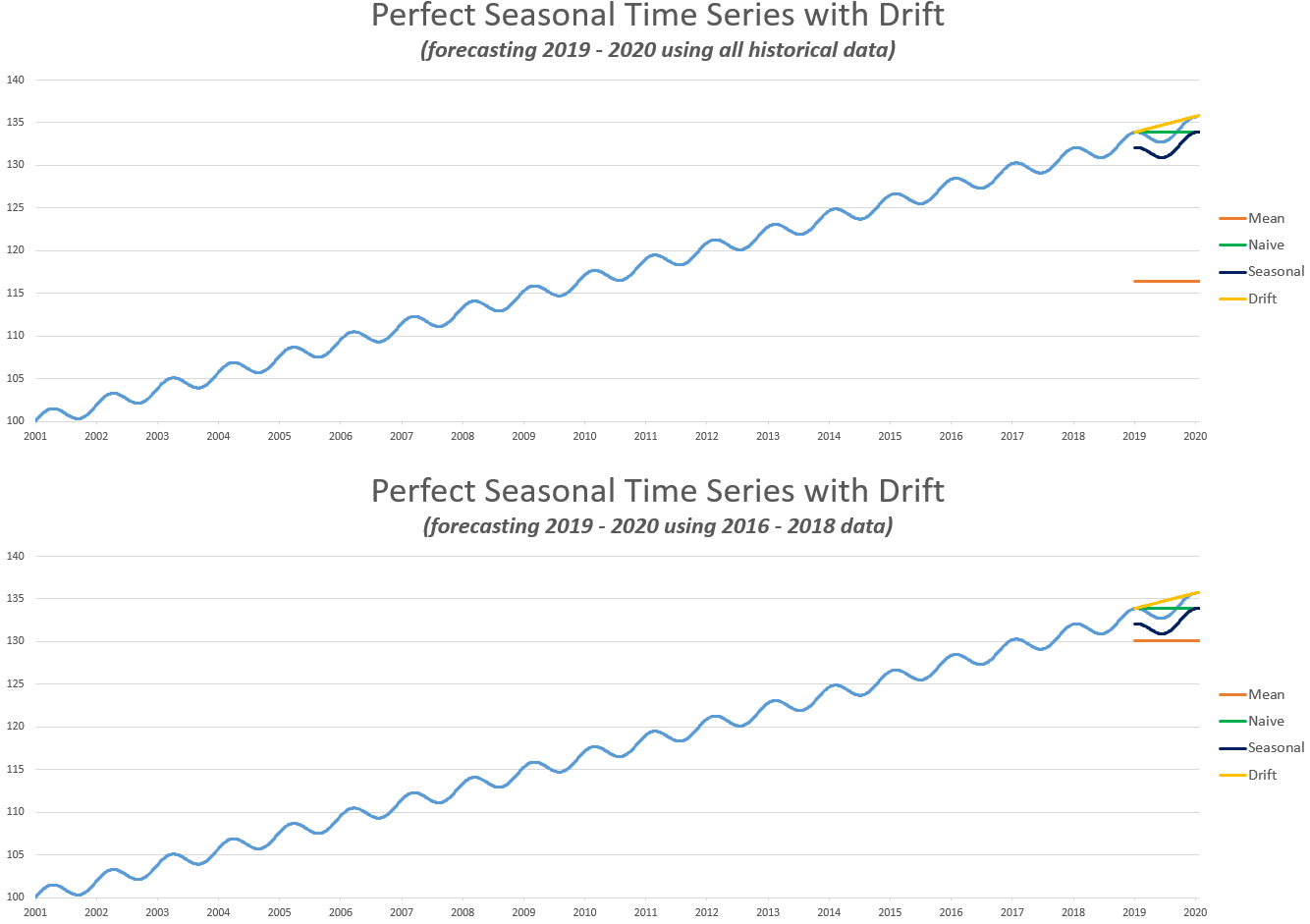
The naïve method works remarkably well for many economic and financial time series because it is the optimal method when data follow a random walk. This is why the naïve method is also referred to as the random walk forecast.

The seasonal naïve method (seasonal method), equation looks more complicated than it really is. For example, with monthly data, the forecast for all future February values is equal to the last observed February value. Similarly, with quarterly data, all future Q1 values is equal to the last observed Q1 value.

The drift method is equivalent to drawing a line between the first and last observation and then extrapolating that line into the future.

In many cases, these simple methods will serve as benchmarks rather than the actual implemented method. This is a critical concept because any new method we develop will be compared to these simple methods. If the new method does not perform better than these simple methods then it will be discarded and the forecaster has to start at square one.

**Figure 2** The mean, naïve, seasonal naïve, and drift methods used to forecast a perfectly seasonal time series with drift for differing historical data timeframes.



§3.2 Transformations and Adjustments

Adjusting historical data can often lead to a simpler forecasting task and this in turn can lead to more accurate forecasts. There are four primary kinds of adjustments:

1. Calendar – removes variation due to simple calendar effects. For example, using average monthly data instead of cumulative monthly data eliminates the variation stemming from differing days in each month.
2. Population – for most data that are affected by population changes, it is best to use per-capita data (per thousand people or per million people) rather than the totals.
3. Inflation – data affected by the value of money is best adjusted for changes in the value of money. For example, instead of using nominal house prices, house prices may be stated in base year dollars. To do this a *relevant* price index is required. In the context of house prices, the core rate of inflation (excludes food and energy components) is a good barometer for inflation relevant to the housing market.

In the y*t* is the nominal price and x*t* is the price at t expressed in base year dollars.

1. Mathematical Transformations – are particularly useful when data shows variation that increases or decreases with the level of the series. A often used mathematical transform is the log transform.

Log transformations are useful because a unit change in the transformed variable translates to a base-*n* change in the untransformed variable. Also, log transforms constrain forecasts to stay positive on the original scale. Other transforms include:

Power Transform:

Box-Cox Transform:

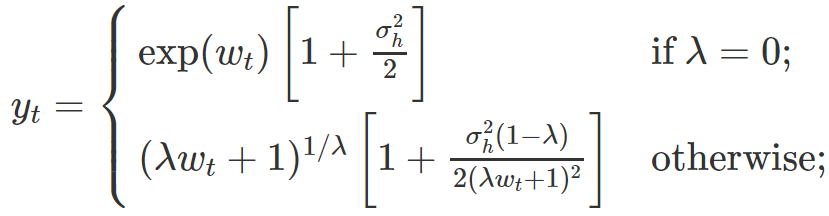
In the Box Cox transform, a good value of λ is one that makes the size of the seasonal variation about the same across the whole series. In R, this is optimal λ can be found using the BoxCox.lambda() function.

The purpose of these adjustments is to simplify patterns in the historical data by removing know sources of variation or by making the pattern more consistent across the whole data set.

Once the original data is transformed, we forecast the transformed data and then reverse the transform (back-transform) to obtain forecasts on the original data scale.

Box-Cox Back-Transform:

***Features of power transformations:***

* If any y*t* is less than zero, power transformation is not possible unless all observations are adjusted by adding a constant to all values.
* Choosing a simple value for λ makes explanations easier.
* Forecasting results are relatively insensitive to the value of λ.
* Often no transformation is needed.
* Transformations sometimes make little difference to the forecasts but have a large effect on prediction intervals.
* For Box Cox transformations, if the distribution on the transformed space is symmetric, then the back-transform usually results in a median forecast instead of a mean forecast. In cases when a median forecast is not acceptable, since means sum but medians do not sum, a bias adjustment is required to attain the mean forecast.

The σ*h*2 is called the h-step forecast variance. The larger the forecast variance, the bigger the difference between the mean and the median. This difference is called the bias and when we use the mean, rather than the median, we say the point forecasts have been bias-adjusted.

***Programmatic Notes***

* *Find the Python equivalent of the R function monthdays()*
* *Find the Python equivalent of the R function BoxCox.lambda()*
* *In R, bias adjustment is not done by default in the forecast package. If you want your forecasts to be means rather than medians, use the argument biasadj=TRUE when you select your Box-Cox transformation parameter.*

§3.3 Residual Diagnostics

Residuals are helpful in checking whether a model has adequately captured the information in the data. Residuals are not however a good way to select the appropriate forecasting method. Residuals from a good forecasting method have the following properties:

1. Uncorrelated - If not then there is uncaptured information.
2. Zero mean - If not, the forecast is biased.
3. Constant variance
4. Normally distributed – this along with property 3 makes prediction interval derivation easier.

In practice: Forecasting Google’s daily closing stock price

At the outset if we use the naïve method, then

There are a few things you can do to analyze the residuals including:

1. Plot the residuals as a time series to observe outliers, stationarity, and skedasticity.
2. Plot the residuals in a histogram to observe normality, skewness, and kurtosis.
3. Plotting the ACF to observe if there is uncaptured information in the residuals.
4. Perform the Portmanteau tests for autocorrelation. We test whether the first h autocorrelations are significantly different from what would be expected from a white noise process.

Box-Pierce test:

h is the maximum lag being considered

T is the number of observations

r*k* is the autocorrelation for lag k

Use h = 10 for non-seasonal data and h = 2m for seasonal data

The test deteriorates for large h; therefore cap h at T/5

Ljung-Box test:

If the autocorrelations did come from a white noise process, then both Q and Q\* would have a χ2 distribution with (h – K) degrees of freedom where K is the number of parameters in the model.

***Programmatic Notes***

* *The R function checkresiduals() offers output that spans many facets of what has been discussed in this section.*