Time Series

Viktor

September 9, 2020

Introduction. Time series are commonly thought as a superposition of several components

- *Trend*: long-term direction of time series

- *Seasonal*: pattern that repeats with known and fixed periodicity

- *Cycle*: pattern that repeats but with unknown und changing periodicity.

- *Autocorrelation*:

- *Error*: unpredictable component of time series.

Many time-series models can be classified as additive, multiplicative, or a combination of these components [1]. The aim to time series modelling is to capture all effects until only error (=noise) is left. More precisely if there ist still come uncaptured trend, seasonal, cylce or autocorrelation then we can be pretty sure that we can improve the model. However, this does not imply that all models that just exhibit unexplained noise perform equally.

Model diagnostics. The autocorrelation function (ACF) is the Pearson correlation of a time dependent random variable Y_t with itself shifted in time by τ.

$$\rho(\tau) = \frac{\sum_{t=\tau+1}^{T}\left(y_t - \bar{y}\right)\left(y_{t-\tau} - \bar{y}\right)}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{1}$$

The autocorrelation function depends on τ. For white noise 95 % of its (τ-dependent) values are within the interval $\pm 2/\sqrt{T}$. Thus the and is used to identify white noise. [3]

Definition 1 (Evaluation metrics). *Let y_t be the observed time series and $\hat{y}_t$ the corresponding predictions of T timesteps.*

- *Mean absolute error:* $\frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t|$

- *Mean squared error:* $\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$.

- *Root mean squared error:* $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$

- *Mean absolute percentage error*

- *Symmetric mean absolute percentage error*

- *Huber loss:*

- *ρ-risk metric:*

Transformations.

- **Calender adjustments.** Sometimes pattern occur that are due to different number of days in a month or if there is some additional known contraint in the data (data points are not generated on sundays). It is much easier to remove this structure by changing to an appropriate time unit than trying to let the model learn this.

- **Correction for known temporal adjustments.** If data is affected by e.g. the population or inflation adjust it in order to account for these effects on the training data level instead of trying to learn these effects. E.g. consider quantities per capita (population) or use inflation corrected data (financial times series).

- **Fourier transformation.** Include the dominant Fourier contributions in order to help the model dealing with periodicity. In practice these components my be found by an FFT and sometimes only the corresponding basis functions are included.

- **Box-Cox transformation** on target values [3],

$$
y_t \mapsto \begin{cases} \ln y_t & \text{if } \lambda = 0 \\ (y_t^\lambda - 1)/\lambda & \text{otherwise.} \end{cases} \tag{2}
$$

  This transformation aims to render the seasonal effects to be about the same across the whole time series, which dictates the choice of λ. In practice forecasting results (mean point estimates) are relatively

insensitive to λ in contrast prediction intervals Once a prediction has been made on the transformed space it needs to be back-transformed into the original scale. The inverse transformation

$$y_t \mapsto \begin{cases} e^{y_t} & \text{if } \lambda = 0 \\ (\lambda y_t + 1)^{\frac{1}{\lambda}} & \text{otherwise,} \end{cases} \tag{3}$$

is sometimes used. However, this point wise inverse of the Box-Cox transformation is *not* the inverse of the underlying forecast distribution [1]. If only the mean as point estimate is of interest then the *biased adjusted* back-transformed mean can be approximated by [2],

$$y_t \mapsto \begin{cases} e^{y_t} \left[1 + \frac{\sigma^2}{2} \right] & \text{if } \lambda = 0 \\ (\lambda y_t + 1)^{\frac{1}{\lambda}} \left[1 + \frac{\sigma^2 (1-\lambda)}{2(\lambda y_t + 1)^{\frac{1}{2}}} \right] & \text{otherwise,} \end{cases} \tag{4}$$

where by assumption y_t is the mean on the transformed scale and σ the variance.

Baseline Methods.

Neural networks.   There are several neural network (NN) based approaches, RNN, LSTM, 1D CNN, deep NN. The basic idea behind deep NN is to translate the problem into a supervised learning problem by predicting y_t with features that are the last l target values $y_{t-1}, y_{t-2}, \ldots y_{t-l}$

Overview of methods

- Regression models

- Exponential smoothing methods

- Arima

- Dynamic Regression

- Hierarchical forecasting

- Vector autoregression

- Neural networks

- survival analysis

---

[1]More precisely, he back transformation (4) is only correct for the median if the transformed forecast distribution is symmetric [3]

References

[1] David J. Hand. Forecasting with exponential smoothing: The state space approach by rob j. hyndman, anne b. koehler, j. keith ord, ralph d. snyder. *International Statistical Review*, 77(2):315–316, 2009.

[2] Robin John Hyndman. The forecast mean after back-transformation, 2014.

[3] Robin John Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition, 2018.