Modeling and Analysis of Time Series Data

Chapter 10: Forecasting

Edward L. Ionides

Outline

Model-based forecasts

- Data, $y^*_{1:N}$, and a model $Y_{1:N+h}$ with joint density $f_{Y_{1:N+h}}(y_{1:N+h}|\theta)$ can be used to **forecast** future values $y_{N+1:N+h}$ up to a **horizon**, h.
- A model-based **probabilistic forecast** of the not-yet-observed values $y_{N+1:N+h}$ is

$$f_{Y_{N+1:N+h}|Y_{1:N}}\big(y_{N+1:N+h}|y^*_{1:N};\hat{\theta}\big), \tag{1}$$

  where $\hat{\theta}$ is a point estimate such as an MLE.
- A model-based **point forecast** of $y_{N+1:N+h}$ is

$$\mathbb{E}\big[Y_{N+1:N+h}\big|Y_{1:N}=y^*_{1:N};\hat{\theta}big\big]. \tag{2}$$

- Point forecasts and probabilistic forecasts have many applications in business and elsewhere.

Evaluating forecasts

- Point forecasts could be evaluated by squared error, absolute error, relative squared error, relative absolute error, etc.
- Probabilistic forecasts are naturally evaluated by the forecast log-density,
$$\log f_{Y_{N+1:N+h}|Y_{1:N}}\big(y_{N+1:N+h}|y^*_{1:N};\hat{\theta}\big), \tag{3}$$
evaluated at the data, $y^*_{N+1:N+h}$, once it is collected.
- Due to time dependence, and limited amounts of data, it can be problematic to evaluate by cross-validation.
- Note that log-likelihood can be written as a sum of one-step forecast log-densities:

$$\log f_{Y_{1:N}}(y^*_{1:N};\theta) = \sum_{n=1}^{N} \log f_{Y_n|Y_{1:n-1}}(y^*_n|Y^*_{1:n-1};\theta) \tag{4}$$

ARIMA forecasting

- predict.Arima() computes the conditional Gaussian distribution for forecasting an ARIMA model.

```
dat <- read.table(file="huron_level.csv",sep=",",header=TRUE)
huron_level <- ts(as.vector(t(dat[,2:13])),start=1860,freq=12)
time <- rep(dat$Year,each=12)+ rep(0:11,nrow(dat))/12
huron_old <- window(huron_level,end=2014.99)
sarma <- arima(huron_old,order=c(1,0,1),
  seasonal=list(order=c(1,0,1),period=12))
f.sarma <- predict(sarma,n.ahead=120)
f.val <- as.vector(f.sarma$pred)
f.se <- as.vector(f.sarma$sd)
f.time <- as.vector(time(f.sarma$pred))
plot(huron_old,xlim=range(time))
lines(f.time,f.val,col="red")
lines(f.time, f.val+1.96*f.se,col="blue")
lines(f.time, f.val-1.96*f.se,col="blue")
lines(time[-seq_along(huron_old)],huron_level[-seq_along(huron_old)]
```

Facebook Prophet

test

Forecasting vs model fitting

test

Further reading

- Section 3.5 of Shumway and Stoffer (2017) covers ARIMA forecasting.
- Hyndman and Khandakar (2008) introduces the `forecast` R package.
- Taylor and Letham (2018) presents the Facebook Prophet forecasting algorithm.

References and Acknowledgements

Hyndman RJ, Khandakar Y (2008). "Automatic time series forecasting: The forecast package for R." *Journal of Statistical Software*, **27**, 1–22.

Shumway RH, Stoffer DS (2017). *Time Series Analysis and its Applications: With R Examples*. 4th edition. Springer.

Taylor SJ, Letham B (2018). "Forecasting at scale." *The American Statistician*, **72**(1), 37–45.

- Compiled on February 18, 2025 using R version 4.4.2.
-
- We acknowledge previous versions of this course.