ISA 444: Business Forecasting

06 - Basic Tools and Goodness of Fit (Cont.)

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Fall 2020

Outline

Quick Refresher based on Last Class

Main Learning Outcomes

- ☑ **Apply transformations to a time series.**
- ☒ **Apply and interpret measures of forecast accuracy.**
- ☒ **Interpret prediction intervals for a simple forecast.**

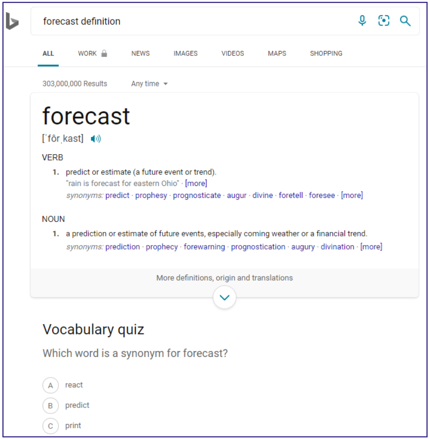
Learning Objectives for Today's Class

Main Learning Outcomes

- **Apply and interpret measures of forecast accuracy.**
- **Interpret prediction intervals for a simple forecast.**

Outline

Recap: Definition of Forecast



The definition of the term "forecast" as obtained from Bing/Merriam-Webster.

A Naive Forecast

- A naïve forecast for an observation, Y_t , is the observation prior, Y_{t-1}.
- For some types of time series (e.g. Random Walks), a naïve forecast is the best possible forecast one can make[1].
- In the case of seasonal data, a naïve forecast could be the observation from the prior period.
    - For example, in the case of monthly data, the naïve forecast for the observation $Y_{Jan2018}$ could be $Y_{Jan2017}$. In this case, we would denote the frequency, m=12, and the naïve forecast for Y_t is the observation m periods prior, or Y_{t-m}.

[1] Slide is from Dr. Allison Jones-Farmer's lecture notes, Miami University, Spring 2020.

Measures of Forecast Accuracy

The measures of accuracy we will discuss all deal with the difference between the actual observed value (Y_t) and the forecasted value (F_t) at time t. In order to measure forecast accuracy, we assume we have m actual values available, thus we have $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+m}$ and forecasts $F_{t+1}, F_{t+2}, \ldots, F_{t+m}$. This is important because we will be averaging the forecast errors over m.[2]

Forecast Error:

$$e_{t+i} = Y_{t+i} - F_{t+i}. \tag{1}$$

[2] Note that your textbook discusses rolling forecast origins. This is important, but we will save this discussion for later in the semester. For now, assume the forecast origin is fixed (i.e., we are only interested in the one-period ahead forecast).

Measures Reflecting "Average" Forecast Performance

Mean Error:

$$ME = \frac{\sum_{i=1}^{m} e_{t+i}}{m}. \tag{2}$$

Mean Percentage Error:

$$MPE = \frac{100}{m} \sum_{i=1}^{m} \frac{e_{t+i}}{Y_{t+i}}. \tag{3}$$

Computing Measures of "Average" Forecast Performance [1]

symbol	date	adjusted	naiveFC	e	PE
AAPL	2020-08-17	114.61			
AAPL	2020-08-18	115.56	114.61	0.96	0.83
AAPL	2020-08-19	115.71	115.56	0.14	0.13
AAPL	2020-08-20	118.28	115.71	2.57	2.17
AAPL	2020-08-21	124.37	118.28	6.10	4.90
AAPL	2020-08-24	125.86	124.37	1.49	1.18
AAPL	2020-08-25	124.82	125.86	-1.03	-0.83
AAPL	2020-08-26	126.52	124.82	1.70	1.34
AAPL	2020-08-27	125.01	126.52	-1.51	-1.21
AAPL	2020-08-28	124.81	125.01	-0.20	-0.16

Based on Approach #1, the ME and MPE are equal to 1.13 and 0.93% , respectively.

Comments:

- We are forecasting the adjusted closing price of the AAPL stock for the next trading day, i.e., we are ignoring non-trading days.
- The naïve forcast is the `lag(1)` of the series; thus, the forecast error is the `diff(1)`.

Computing Measures of “Average” Forecast Performance [2]

```r
pacman::p_load(tidyquant, magrittr, fpp2)

aapl = tq_get("AAPL", from = "2020-08-17", to = "2020-08-29") %>%
  select(symbol, date, adjusted)
naiveFC = snaive(aapl$adjusted) %>% .[['fitted']] #snaive from fpp2

e = aapl$adjusted - naiveFC
ME = mean(e, na.rm = T)
MPE = 100*mean(e/aapl$adjusted, na.rm = T)

cat(paste0("Based on Approach #2, the ME and MPE are equal to ", round(ME, 2)
           "%, respectively."))
```

Based on Approach #2, the ME and MPE are equal to 1.13 and 0.93%, respectively.

Measures Reflecting “Variablity” in Forecast Performance

Absolute Forecast Error:

$$|e_{t+i}| = |Y_{t+i} - F_{t+i}|. \tag{4}$$

Squared Forecast Error:

$$(e_{t+i})^2 = (Y_{t+i} - F_{t+i})^2. \tag{5}$$

Mean Absolute Error:

$$MAE = \frac{\sum_{i=1}^{m} |e_{t+i}|}{m}. \tag{6}$$

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} (e_{t+i})^2}{m}}. \tag{7}$$

Measures Reflecting “Relative” Forecast Performance

Mean Absolute Percentage Error:

$$MAPE = \frac{100}{m} \sum_{i=1}^{m} \frac{|e_{t+i}|}{m}. \tag{8}$$

Relative Mean Absolute Error:

$$RelMAE = \frac{\sum_{i=1}^{m} |e_{t+i}|}{\sum_{i=1}^{m} |Y_{t+i} - Y_{t+i-1}|}. \tag{9}$$

Thiel’s U:

$$U = \sqrt{\frac{\sum_{i=1}^{m} (e_{t+i})^2}{\sum_{i=1}^{m} (Y_{t+i} - Y_{t+i-1})^2}}. \tag{10}$$

An Overview of Computing these Measures in R [1]

```
pacman::p_load(tidyquant, magrittr, fpp2, xtable)

aapl = tq_get("AAPL", from = "2020-08-17", to = "2020-08-29") %>%
  select(symbol, date, adjusted)
naiveFC = lag(aapl$adjusted)

e = aapl$adjusted - naiveFC
ME = mean(e, na.rm=T)
RMSE = mean(e^2, na.rm=T) %>% sqrt()
MAE = abs(e) %>% mean(na.rm=T)
MPE = 100 * mean(e/aapl$adjusted, na.rm=T)
MAPE= 100 * mean(abs(e)/aapl$adjusted, na.rm=T)
```

An Overview of Computing these Measures in R [2]

```r
E = c(ME, RMSE, MAE, MPE, MAPE)
names(E) = c("ME", "RMSE", "MAE", "MPE", "MAPE")
round(E, 2) %>% print()
```

```
##   ME RMSE  MAE  MPE MAPE
## 1.13 2.43 1.74 0.93 1.42
```

```r
# Alternatively, we could have just computed it using the fpp2 package
accuracy(naiveFC, aapl$adjusted) %>% round(2)
```

```
##             ME RMSE  MAE  MPE MAPE
## Test set 1.13 2.43 1.74 0.93 1.42
```

Outline

Point vs Interval Forecasts

- **Point Forecasts:** future observations for which we report a single forecast observation.

- **Interval Forecast:** a range of values that are reported to forecast an outcome.

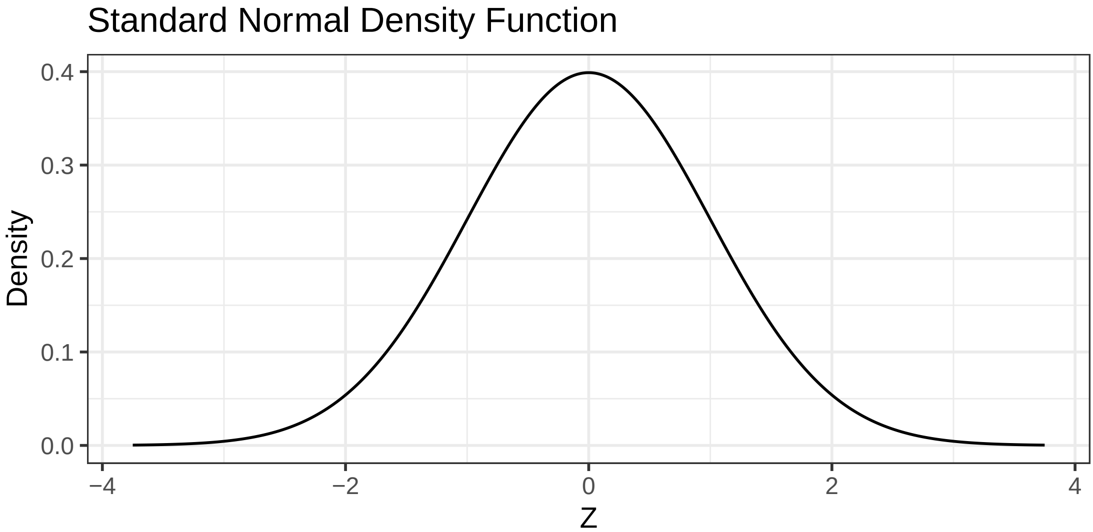
If we assume the forecast errors follow a Normal Distribution, an approximate $100(1-\alpha)$ prediction interval can be computed as follows:

$$\hat{F}_t \pm Z_{\frac{\alpha}{2}} * RMSE, \tag{11}$$

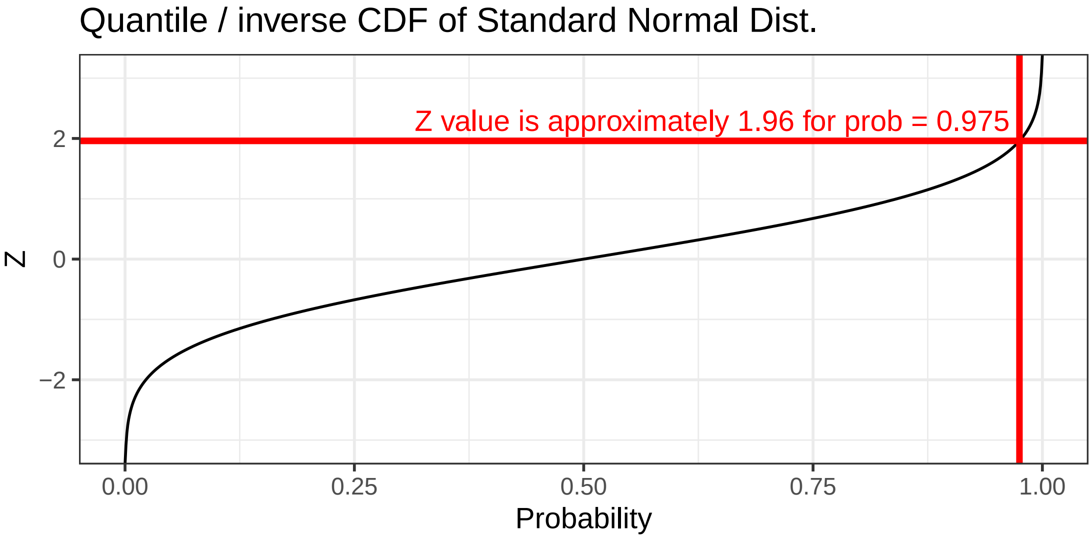
where:

- $\hat{F}_t$ forecast at time t.
- The RMSE can be used as an estimate of the standard deviation of the forecast errors.
- Z is the quantile corresponding to $100(1 - \frac{\alpha}{2})$.

Recall: Standard Normal Distribution [1]



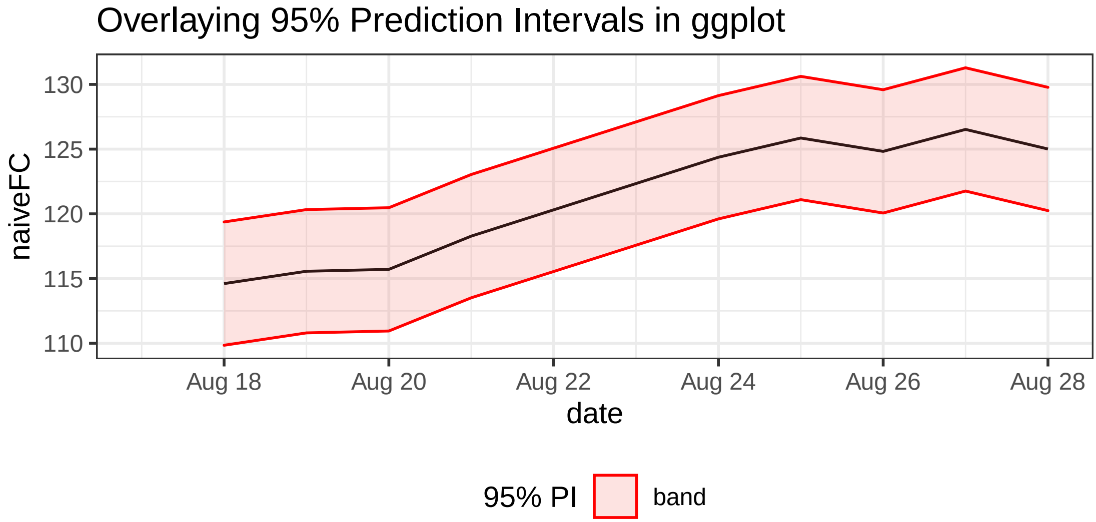
Recall: Standard Normal Distribution [2]



Prediction Intervals for the $AAPL Data [1]

```r
PInormU = naiveFC + abs(qnorm(0.975))*RMSE
PInormL = naiveFC - abs(qnorm(0.975))*RMSE
dfNaive = data.frame(date = aapl$date, Ft = naiveFC, PInormU, PInormL)
dfNaive %>% ggplot(aes(x= date, y = naiveFC)) + geom_line() +
  theme_bw() +
  labs(title ="Overlaying 95% Prediction Intervals in ggplot",
       fill = "95% PI") +
    geom_ribbon(aes(ymin = PInormL , ymax = PInormU, fill = "band"),
                alpha = 0.2, color = "red") +
  theme(legend.position= "bottom")
```

Prediction Intervals for the $AAPL Data [2]



Outline

Summary of Main Points

Main Learning Outcomes

- **Apply and interpret measures of forecast accuracy.**
- **Interpret prediction intervals for a simple forecast.**

Things to Do

- Thoroughly read all of Chapter 2 of our book.
- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- Recreate the plots in Slides 15-17 for the RSCCASN Data. Note that the period, $m = 12$.
- If you are interested in additional practice problems, please consider the following problems from your textbook. To access these datasets, please click here.
    - For the Means approaches in Example 2.7 (P.49), use R to compute the 7 error measures for the four forecasting approaches.
    - Exercise 2.12 and compute the forecast errors for the naive forecast.

Graded Assignment 04: Evaluating your Retention/Focus

Please go to Canvas (click here) and answer the questions. **Due Sept. 3, 2020 [2:50 PM, Ohio local time].**

What/Why/Prep? The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 04. In order to prepare for this, you should have either actively attended class and/or watched the recording from WebEx. Furthermore, you should have thoroughly read up to the end of Chapter 2 from your textbook.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment (i.e. once you start the assignment you will have 10-15 minutes to complete the one question).
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the Remote Desktop)

Graded Assignment 05: Evaluating your Retention/Focus

Please go to Canvas (click here) and answer the two questions. **Due Sept. 7, 2020 [2:50 PM, Ohio local time].**

What/Why/Prep? The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 05. In order to prepare for this, you should have either actively attended class and/or watched the recording from WebEx. Furthermore, you should have thoroughly read up to the end of Chapter 2 from your textbook.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment (i.e. once you start the assignment you will have 10-15 minutes to complete the one question).
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the Remote Desktop)

ISA 444: Business Forecasting

06 - Basic Tools and Goodness of Fit (Cont.)

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Fall 2020