

## 5. The forecaster's toolbox

### 5.5 Distributional forecasts

[OTexts.org/fpp3/](http://OTexts.org/fpp3/)

Rob J Hyndman  
George Athanasopoulos

# FORECASTING

## PRINCIPLES AND PRACTICE

A comprehensive introduction to the latest forecasting methods using R. Learn to improve your forecast accuracy using dozens of real data examples.



3RD EDITION

 **OTexts**  
OPEN ACCESS TEXTS

# Forecast distributions

- A forecast  $\hat{y}_{T+h|T}$  is (usually) the mean of the conditional distribution  $y_{T+h} \mid y_1, \dots, y_T$ .
- Most time series models produce normally distributed forecasts.
- The forecast distribution describes the probability of observing any future value.

# Forecast distributions

Assuming residuals are normal, uncorrelated,  $\text{sd} = \hat{\sigma}$ :

**Mean:**  $y_{T+h|T} \sim N(\bar{y}, (1 + 1/T)\hat{\sigma}^2)$

**Naïve:**  $y_{T+h|T} \sim N(y_T, h\hat{\sigma}^2)$

**Seasonal naïve:**  $y_{T+h|T} \sim N(y_{T+h-m(k+1)}, (k+1)\hat{\sigma}^2)$

**Drift:**  $y_{T+h|T} \sim N(y_T + \frac{h}{T-1}(y_T - y_1), h\frac{T+h}{T}\hat{\sigma}^2)$

where  $k$  is the integer part of  $(h - 1)/m$ .

Note that when  $h = 1$  and  $T$  is large, these all give the same approximate forecast variance:  $\hat{\sigma}^2$ .

# Prediction intervals

- A prediction interval gives a region within which we expect  $y_{T+h}$  to lie with a specified probability.
- Assuming forecast errors are normally distributed, then a 95% PI is

$$\hat{y}_{T+h|T} \pm 1.96\hat{\sigma}_h$$

where  $\hat{\sigma}_h$  is the st dev of the  $h$ -step distribution.

- When  $h = 1$ ,  $\hat{\sigma}_h$  can be estimated from the residuals.

# Prediction intervals

```
aus_production |>
  filter(!is.na(Bricks)) |>
  model(Seasonal_naive = SNAIVE(Bricks)) |>
  forecast(h = "5 years")
```

```
## # A tibble: 20 x 4 [1Q]
```

```
## # Key:           .model [1]
```

##	.model	Quarter	Bricks	.mean
##	<chr>	<qtr>	<dist>	<dbl>
##	1 Seasonal_naive	2005 Q3	N(428, 2336)	428
##	2 Seasonal_naive	2005 Q4	N(397, 2336)	397
##	3 Seasonal_naive	2006 Q1	N(355, 2336)	355
##	4 Seasonal_naive	2006 Q2	N(435, 2336)	435
##	5 Seasonal_naive	2006 Q3	N(428, 4672)	428
##	6 Seasonal_naive	2006 Q4	N(397, 4672)	397
##	7 Seasonal_naive	2007 Q1	N(355, 4672)	355

# Prediction intervals

```
aus_production |>
  filter(!is.na(Bricks)) |>
  model(Seasonal_naive = SNAIVE(Bricks)) |>
  forecast(h = "5 years") |>
  hilo(level = 95)
```

```
## # A tsibble: 20 x 5 [1Q]
```

```
## # Key:           .model [1]
```

##	.model	Quarter	Bricks	.mean	`95%`
##	<chr>	<qtr>	<dist>	<dbl>	<hilo>
##	1 Seasonal_naive	2005 Q3	N(428, 2336)	428	[333, 523]95
##	2 Seasonal_naive	2005 Q4	N(397, 2336)	397	[302, 492]95
##	3 Seasonal_naive	2006 Q1	N(355, 2336)	355	[260, 450]95
##	4 Seasonal_naive	2006 Q2	N(435, 2336)	435	[340, 530]95
##	5 Seasonal_naive	2006 Q3	N(428, 4672)	428	[294, 562]95
##	6 Seasonal_naive	2006 Q4	N(397, 4672)	397	[263, 531]95

# Prediction intervals

```
aus_production |>
  filter(!is.na(Bricks)) |>
  model(Seasonal_naive = SNAIVE(Bricks)) |>
  forecast(h = "5 years") |>
  hilo(level = 95) |>
  unpack_hilo("95%")
```

```
## # A tsibble: 20 x 6 [1Q]
```

```
## # Key:           .model [1]
```

##	.model	Quarter	Bricks	.mean	`95%_lower`	`95%_upper`
##	<chr>	<qtr>	<dist>	<dbl>	<dbl>	<dbl>
##	1 Seasonal_naive	2005 Q3	N(428, 2336)	428	333.	523.
##	2 Seasonal_naive	2005 Q4	N(397, 2336)	397	302.	492.
##	3 Seasonal_naive	2006 Q1	N(355, 2336)	355	260.	450.
##	4 Seasonal_naive	2006 Q2	N(435, 2336)	435	340.	530.
##	5 Seasonal_naive	2006 Q3	N(428, 4672)	428	294.	562.

# Prediction intervals

- Point forecasts are often useless without a measure of uncertainty (such as prediction intervals).
- Prediction intervals require a stochastic model (with random errors, etc).
- For most models, prediction intervals get wider as the forecast horizon increases.
- Use `level` argument to control coverage.
- Check residual assumptions before believing them.
- Prediction intervals are usually too narrow due to unaccounted uncertainty.