

## 5. The forecaster's toolbox

### 5.4 Residual diagnostics

[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.

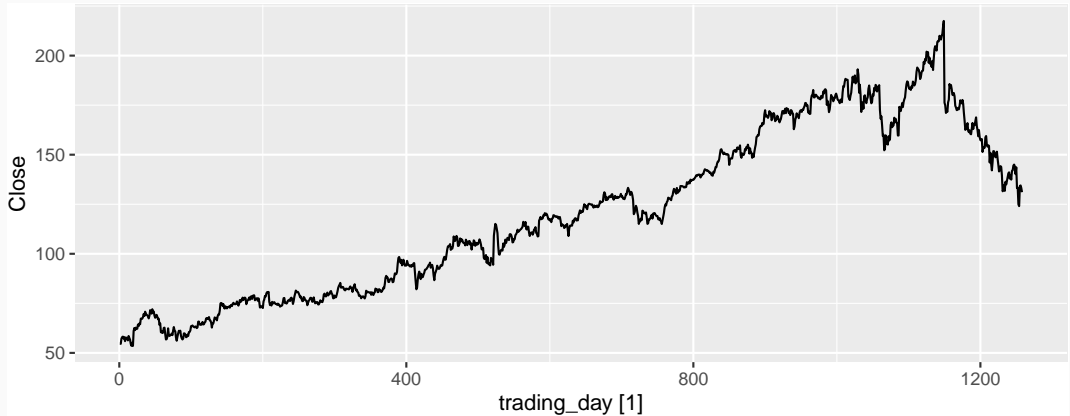


3RD EDITION

 **OTexts**  
OPEN TEXTS FOR PRACTICE

# Facebook closing stock price

```
fb_stock |> autoplot(Close)
```



# Facebook closing stock price

```
fit <- fb_stock |> model(NAIVE(Close))  
augment(fit)
```

```
## # A tsibble: 1,258 x 7 [1]  
## # Key:      Symbol, .model [1]  
##   Symbol .model      trading_day Close .fitted .resid .innov  
##   <chr>  <chr>          <int> <dbl>  <dbl>  <dbl>  <dbl>  
## 1 FB    NAIVE(Close)      1  54.7   NA     NA     NA  
## 2 FB    NAIVE(Close)      2  54.6   54.7  -0.150 -0.150  
## 3 FB    NAIVE(Close)      3  57.2   54.6   2.64   2.64  
## 4 FB    NAIVE(Close)      4  57.9   57.2   0.720  0.720  
## 5 FB    NAIVE(Close)      5  58.2   57.9   0.310  0.310  
## 6 FB    NAIVE(Close)      6  57.2   58.2  -1.01  -1.01  
## 7 FB    NAIVE(Close)      7  57.9   57.2   0.720  0.720  
## 8 FB    NAIVE(Close)      8  55.9   57.9  -2.03  -2.03  
## 9 FB    NAIVE(Close)      9  57.7   55.9   1.83   1.83  
## 10 FB   NAIVE(Close)     10  57.6   57.7  -0.140 -0.140  
## # with 1,248 more rows
```

# Facebook closing stock price

```
fit <- fb_stock |> model(NAIVE(Close))  
augment(fit)
```

```
## # A tsibble: 1,258 x 7 [1]  
## # Key:      Symbol, .model [1]  
##   Symbol .model      trading_day Close .fitted .resid .innov  
##   <chr>  <chr>          <int>  <dbl>  <dbl>  <dbl>  <dbl>  
## 1 FB     NAIVE(Close)      1    54.7    NA    NA     NA  
## 2 FB     NAIVE(Close)      2    54.6    54.7 -0.150 -0.150  
## 3 FB     NAIVE(Close)      3    57.2    54.6  2.64   2.64  
## 4 FB     NAIVE(Close)      4    57.9    57.2  0.720  0.720  
## 5 FB     NAIVE(Close)      5    58.2    57.9  0.310  0.310  
## 6 FB     NAIVE(Close)      6    57.2    58.2 -1.01  -1.01  
## 7 FB     NAIVE(Close)      7    57.9    57.2  0.720  0.720  
## 8 FB     NAIVE(Close)      8    55.9    57.9 -2.03  -2.03  
## 9 FB     NAIVE(Close)      9    57.7    55.9  1.83   1.83  
## 10 FB    NAIVE(Close)     10    57.6    57.7 -0.140 -0.140
```

 $\hat{y}_{t|t-1}$  $e_t$ 

## Naïve forecasts:

$$\hat{y}_{t|t-1} = y_{t-1}$$

$$e_t = y_t - \hat{y}_{t|t-1} = y_t - y_{t-1}$$

```
## # with 1,248 more rows
```

# Facebook closing stock price

```
augment(fit) |>  
  ggplot(aes(x = trading_day)) +  
  geom_line(aes(y = Close, colour = "Data")) +  
  geom_line(aes(y = .fitted, colour = "Fitted"))
```



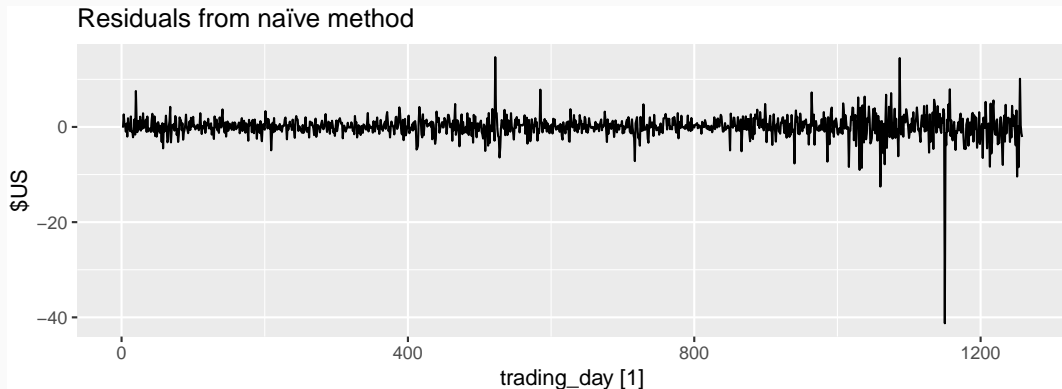
# Facebook closing stock price

```
augment(fit) |>  
  filter(trading_day > 1100) |>  
  ggplot(aes(x = trading_day)) +  
  geom_line(aes(y = Close, colour = "Data")) +  
  geom_line(aes(y = .fitted, colour = "Fitted"))
```



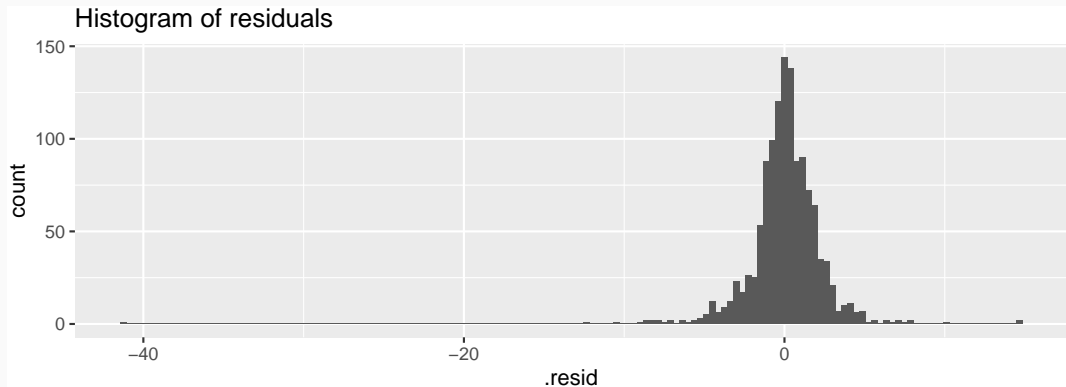
# Facebook closing stock price

```
augment(fit) |>  
  autoplot(.resid) +  
  labs(y = "$US",  
       title = "Residuals from naïve method")
```



# Facebook closing stock price

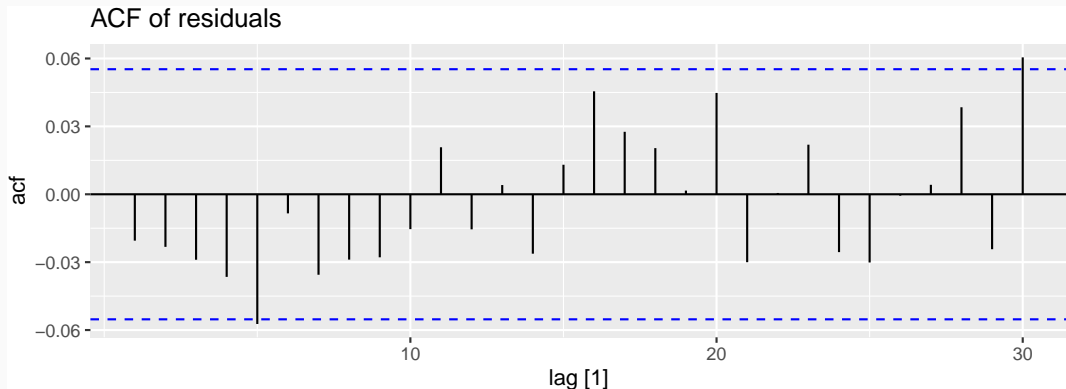
```
augment(fit) |>  
  ggplot(aes(x = .resid)) +  
  geom_histogram(bins = 150) +  
  labs(title = "Histogram of residuals")
```





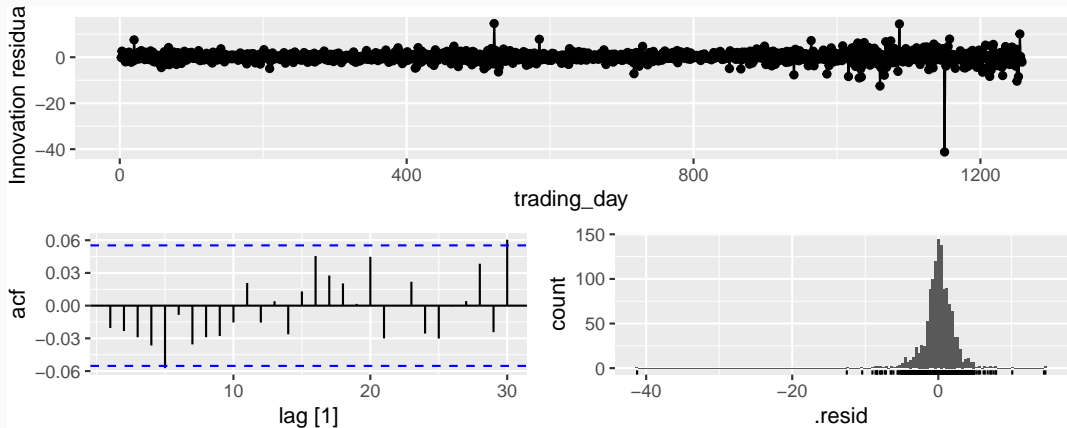
# Facebook closing stock price

```
augment(fit) |>  
  ACF(.resid) |>  
  autoplot() + labs(title = "ACF of residuals")
```



# gg\_tsresiduals() function

```
gg_tsresiduals(fit)
```



## ACF of residuals

- We assume that the residuals are white noise (uncorrelated, mean zero, constant variance). If they aren't, then there is information left in the residuals that should be used in computing forecasts.
- So a standard residual diagnostic is to check the ACF of the residuals of a forecasting method.
- We *expect* these to look like white noise.

# Portmanteau tests

$r_k$  = autocorrelation of residual at lag  $k$

Consider a *whole set* of  $r_k$  values, and develop a test to see whether the set is significantly different from a zero set.

# Portmanteau tests

$r_k$  = autocorrelation of residual at lag  $k$

Consider a *whole set* of  $r_k$  values, and develop a test to see whether the set is significantly different from a zero set.

## Box-Pierce test

$$Q = T \sum_{k=1}^{\ell} r_k^2$$

where  $\ell$  is max lag being considered and  $T$  is number of observations.

- If each  $r_k$  close to zero,  $Q$  will be **small**.
- If some  $r_k$  values large (positive or negative),  $Q$  will be **large**.

# Portmanteau tests

$r_k$  = autocorrelation of residual at lag  $k$

Consider a *whole set* of  $r_k$  values, and develop a test to see whether the set is significantly different from a zero set.

## Ljung-Box test

$$Q^* = T(T+2) \sum_{k=1}^{\ell} (T-k)^{-1} r_k^2$$

where  $\ell$  is max lag being considered and  $T$  is number of observations.

- My preferences:  $\ell = 10$  for non-seasonal data,  $h = 2m$  for seasonal data (where  $m$  is seasonal period).
- Better performance, especially in small samples.

# Portmanteau tests

- If data are WN,  $Q^*$  has  $\chi^2$  distribution with  $\ell$  degrees of freedom.
- $\text{lag} = \ell$

```
augment(fit) |>  
  features(.resid, ljung_box, lag = 10)
```

```
## # A tibble: 1 x 4  
##   Symbol .model      lb_stat lb_pvalue  
##   <chr>   <chr>      <dbl>    <dbl>  
## 1 FB     NAIVE(Close)    12.1     0.276
```