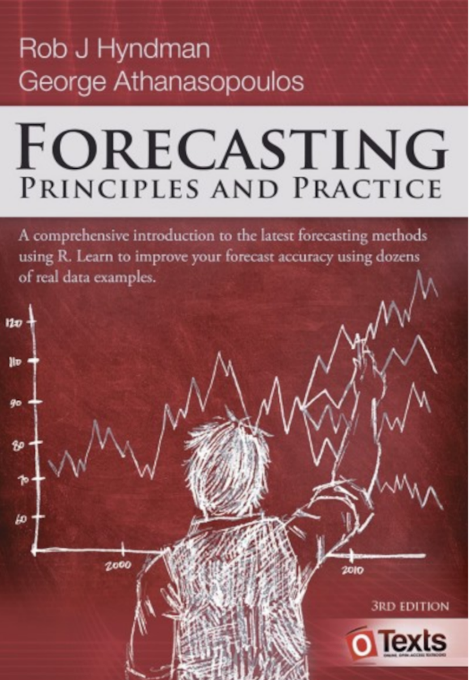


9. ARIMA models

9.9 Seasonal ARIMA models

OTexts.org/fpp3/

Seasonal ARIMA models

ARIMA	$\underbrace{(p, d, q)}$	$\underbrace{(P, D, Q)_m}$
	$\uparrow$ Non-seasonal part of the model	$\uparrow$ Seasonal part of of the model

where m = number of observations per year.

Seasonal ARIMA models

E.g., ARIMA$(1, 1, 1)(1, 1, 1)_4$ model (without constant)

Seasonal ARIMA models

E.g., ARIMA(1, 1, 1)(1, 1, 1)$_4$ model (without constant)

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t \;=\; (1 + \theta_1 B)(1 + \Theta_1 B^4)\varepsilon_t.$$

Seasonal ARIMA models

E.g., ARIMA$(1, 1, 1)(1, 1, 1)_4$ model (without constant)

$$\underbrace{(1-\phi_1 B)}_{\text{Non-seasonal AR(1)}}\underbrace{(1-\Phi_1 B^4)}_{\text{Seasonal AR(1)}}\underbrace{(1-B)}_{\text{Non-seasonal difference}}\underbrace{(1-B^4)}_{\text{Seasonal difference}}y_t = \underbrace{(1+\theta_1 B)}_{\text{Non-seasonal MA(1)}}\underbrace{(1+\Theta_1 B^4)}_{\text{Seasonal MA(1)}}\varepsilon_t.$$

Seasonal ARIMA models

E.g., ARIMA(1, 1, 1)(1, 1, 1)$_4$ model (without constant)

$$(1-\phi_1 B)(1-\Phi_1 B^4)(1-B)(1-B^4)y_t \;=\; (1+\theta_1 B)(1+\Theta_1 B^4)\varepsilon_t.$$

All the factors can be multiplied out and the general model written as follows:

$$\begin{aligned} y_t = {} & (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + (1+\Phi_1)y_{t-4} \\ & -(1+\phi_1+\Phi_1+\phi_1\Phi_1)y_{t-5} + (\phi_1+\phi_1\Phi_1)y_{t-6} \\ & -\Phi_1 y_{t-8} + (\Phi_1+\phi_1\Phi_1)y_{t-9} - \phi_1\Phi_1 y_{t-10} \\ & +\varepsilon_t + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-4} + \theta_1\Theta_1\varepsilon_{t-5}. \end{aligned}$$

Seasonal ARIMA models

The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF.

ARIMA(0,0,0)(0,0,1)$_{12}$ will show:

- a spike at lag 12 in the ACF but no other significant spikes.
- The PACF will show exponential decay in the seasonal lags; that is, at lags 12, 24, 36, ...

ARIMA(0,0,0)(1,0,0)$_{12}$ will show:

- exponential decay in the seasonal lags of the ACF
- a single significant spike at lag 12 in the PACF.

How does ARIMA() work for seasonal models?

A seasonal ARIMA process

$$\Phi(B)\phi(B)(1-B)^d(1-B)^D y_t = c + \Theta(B)\theta(B)\varepsilon_t$$

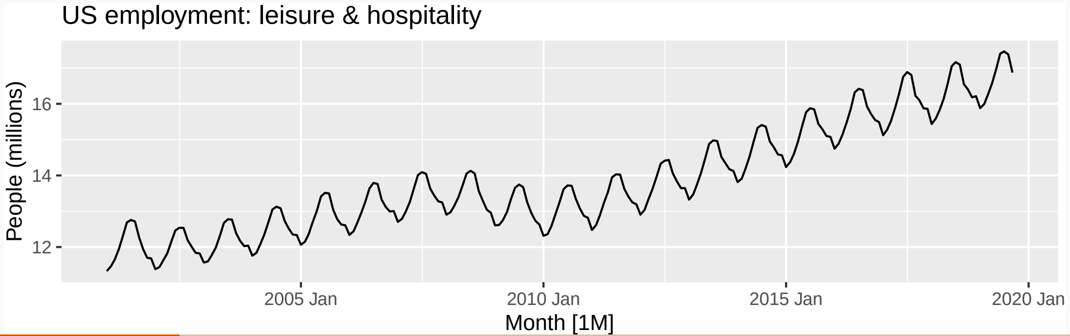
Need to select appropriate orders: d, D, p, q, P, Q, and whether to include the intercept c.

Hyndman and Khandakar (JSS, 2008) algorithm:

- Select no. differences d via KPSS test and D using seasonal strength.
- Select p, q, P, Q and c by minimising AICc.
- Use stepwise search to traverse model space.

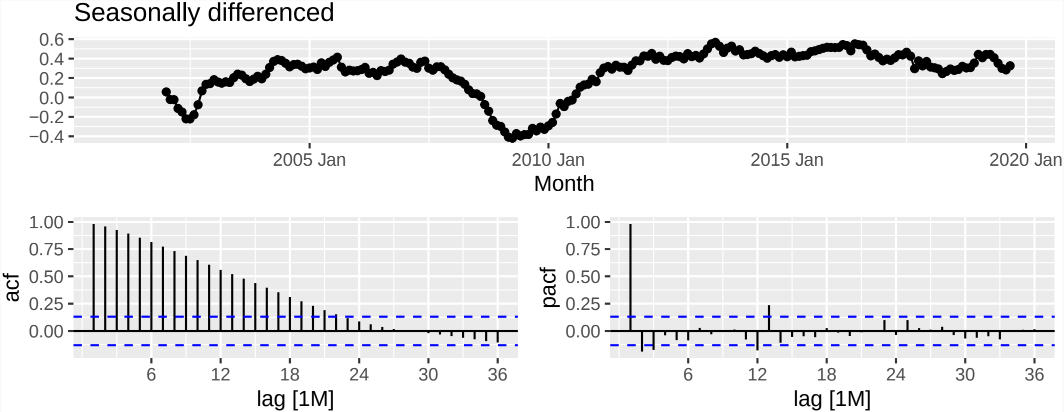
US leisure employment

```r
leisure <- us_employment |>
  filter(Title == "Leisure and Hospitality", year(Month) > 2000) |>
  mutate(Employed = Employed / 1000) |>
  select(Month, Employed)
autoplot(leisure, Employed) +
  labs(title = "US employment: leisure & hospitality", y = "People (millions)")
```



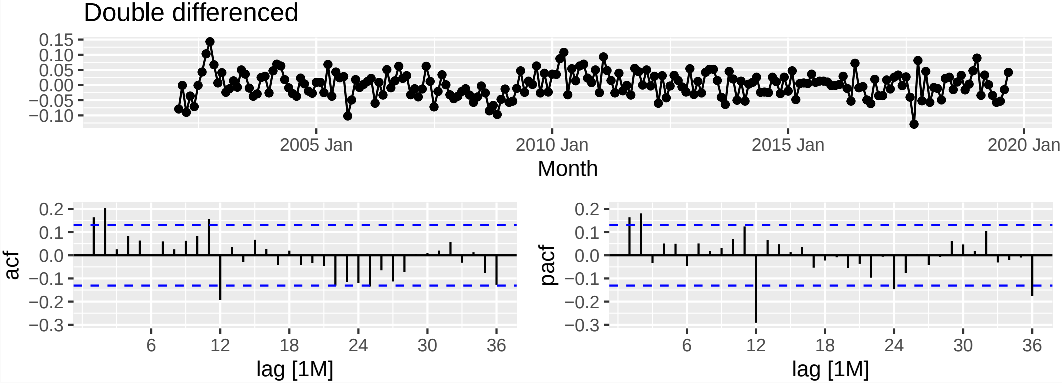
US leisure employment

```r
leisure |>
  gg_tsdisplay(difference(Employed, 12), plot_type = "partial", lag = 36) +
  labs(title = "Seasonally differenced", y = "")
```



US leisure employment

```r
leisure |>
  gg_tsdisplay(difference(Employed, 12) |> difference(),
    plot_type = "partial", lag = 36) +
  labs(title = "Double differenced", y = "")
```



US leisure employment

```r
fit <- leisure |>
  model(arima012011 = ARIMA(Employed ~ pdq(0, 1, 2) + PDQ(0, 1, 1)),
        arima210011 = ARIMA(Employed ~ pdq(2, 1, 0) + PDQ(0, 1, 1)),
        auto = ARIMA(Employed, stepwise = FALSE, approx = FALSE))
fit |>
  pivot_longer(everything(),
    names_to = "Model name",
    values_to = "Orders")
```

```
## # A mable: 3 x 2
## # Key:     Model name [3]
##   `Model name`                    Orders
##   <chr>                          <model>
## 1 arima012011  <ARIMA(0,1,2)(0,1,1)[12]>
## 2 arima210011  <ARIMA(2,1,0)(0,1,1)[12]>
## 3 auto         <ARIMA(2,1,0)(1,1,1)[12]>
```

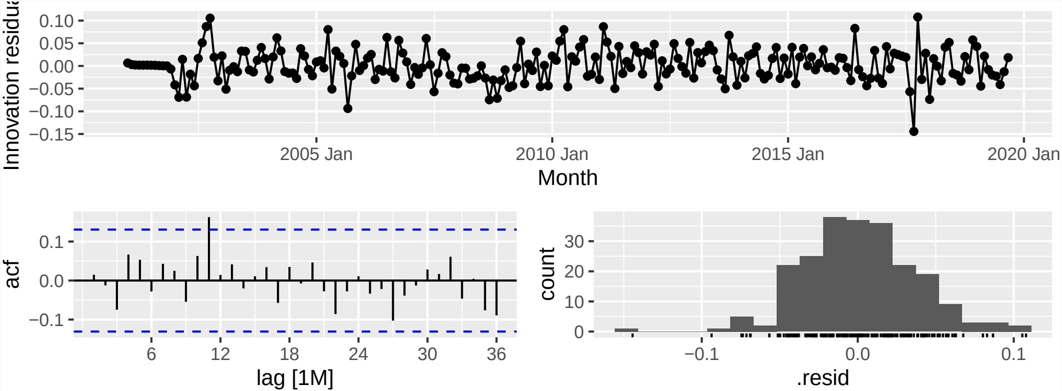
US leisure employment

```r
glance(fit) |>
  arrange(AICc) |>
  select(.model:BIC)
```

```
## # A tibble: 3 x 6
##   .model        sigma2 log_lik   AIC  AICc   BIC
##   <chr>          <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 auto         0.00142    395. -780. -780. -763.
## 2 arima210011 0.00145    392. -776. -776. -763.
## 3 arima012011 0.00146    391. -775. -775. -761.
```

US leisure employment

```
fit |>
  select(auto) |>
  gg_tsresiduals(lag = 36)
```



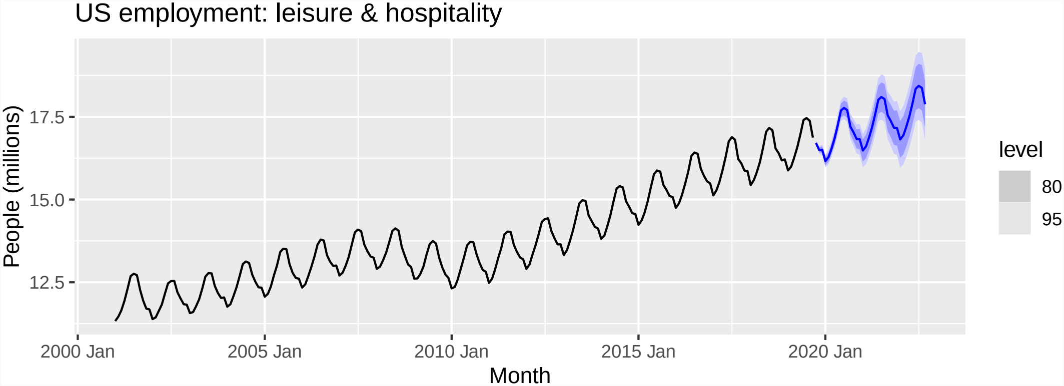
US leisure employment

```r
augment(fit) |>
  filter(.model == "auto") |>
  features(.innov, ljung_box, lag = 24, dof = 4)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 auto      16.6     0.680
```

US leisure employment

```r
forecast(fit, h = 36) |>
  filter(.model == "auto") |>
  autoplot(leisure) +
  labs(title = "US employment: leisure & hospitality", y = "People (millions)")
```

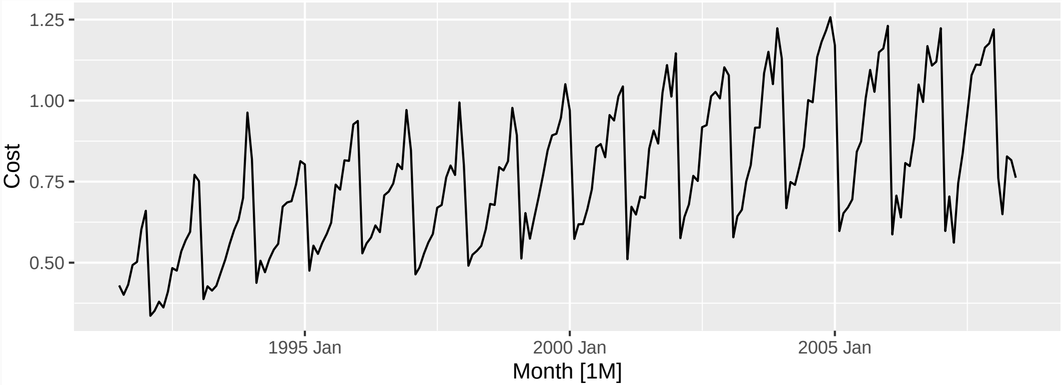


Corticosteroid drug sales

```r
h02 <- PBS |>
  filter(ATC2 == "H02") |>
  summarise(Cost = sum(Cost) / 1e6)
```

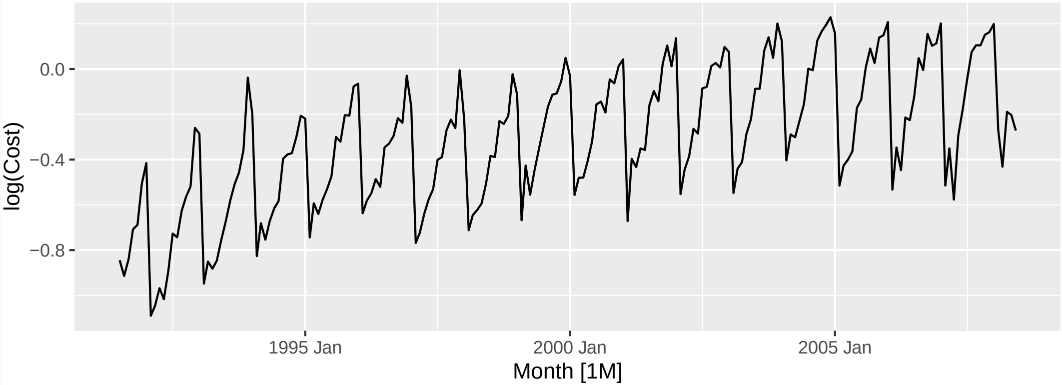
Corticosteroid drug sales

```r
h02 |> autoplot(
  Cost
)
```



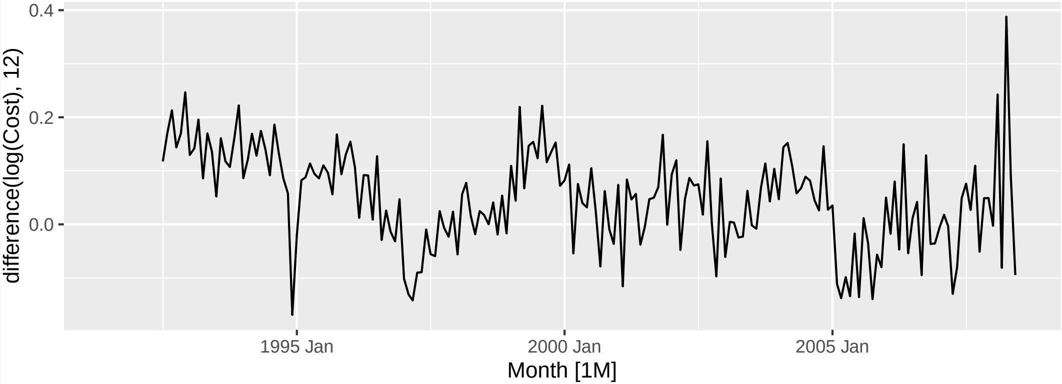
Corticosteroid drug sales

```r
h02 |> autoplot(
  log(Cost)
)
```



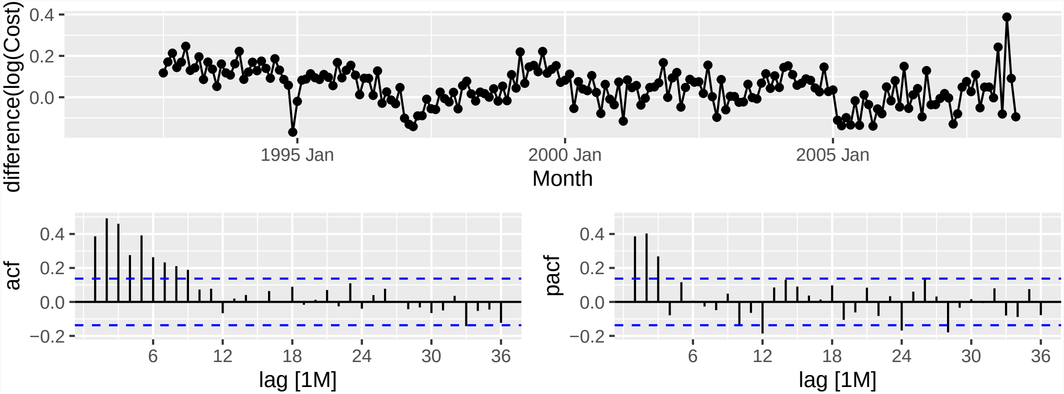
Corticosteroid drug sales

```
h02 |> autoplot(
  log(Cost) |> difference(12)
)
```



Corticosteroid drug sales

```r
h02 |> gg_tsdisplay(difference(log(Cost), 12),
  lag_max = 36, plot_type = "partial"
)
```



Corticosteroid drug sales

- Choose $D = 1$ and $d = 0$.
- Spikes in PACF at lags 12 and 24 suggest seasonal AR(2) term.
- Spikes in PACF suggests possible non-seasonal AR(3) term.
- Initial candidate model: ARIMA$(3,0,0)(2,1,0)_{12}$.

Corticosteroid drug sales

.model	AICc
ARIMA(3,0,1)(0,1,2)[12]	-485
ARIMA(3,0,1)(1,1,1)[12]	-484
ARIMA(3,0,1)(0,1,1)[12]	-484
ARIMA(3,0,1)(2,1,0)[12]	-476
ARIMA(3,0,0)(2,1,0)[12]	-475
ARIMA(3,0,2)(2,1,0)[12]	-475
ARIMA(3,0,1)(1,1,0)[12]	-463

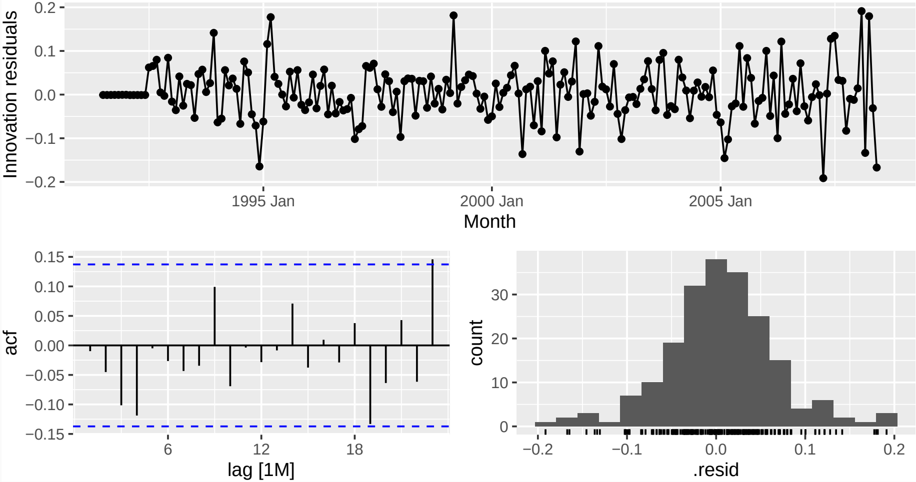
Corticosteroid drug sales

```r
fit <- h02 |>
  model(best = ARIMA(log(Cost) ~ 0 + pdq(3, 0, 1) + PDQ(0, 1, 2)))
report(fit)
```

```
## Series: Cost
## Model: ARIMA(3,0,1)(0,1,2)[12]
## Transformation: log(Cost)
##
## Coefficients:
##          ar1     ar2     ar3    ma1     sma1     sma2
##       -0.160  0.5481  0.5678  0.383  -0.5222  -0.1768
## s.e.   0.164  0.0878  0.0942  0.190   0.0861   0.0872
##
## sigma^2 estimated as 0.004278:  log likelihood=250
## AIC=-486   AICc=-485   BIC=-463
```

Corticosteroid drug sales

```r
gg_tsresiduals(fit)
```



Corticosteroid drug sales

```r
augment(fit) |>
  features(.innov, ljung_box, lag = 36, dof = 6)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 best      50.7    0.0104
```

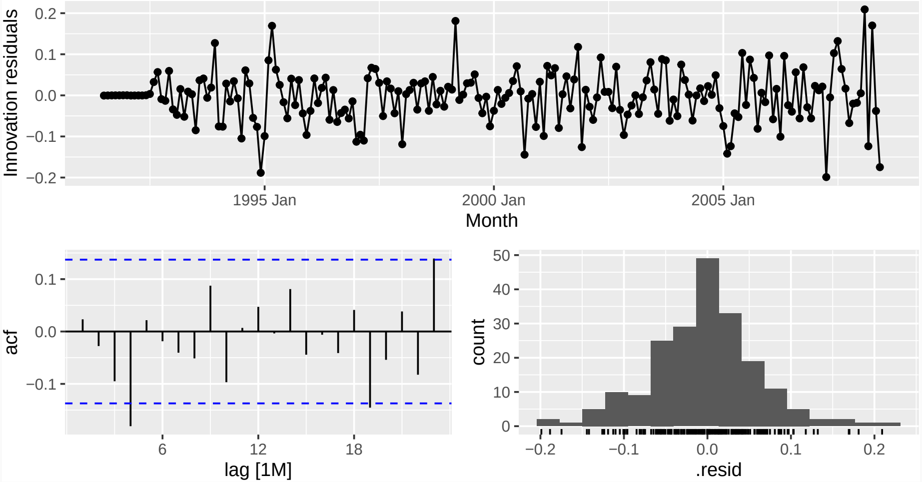
Corticosteroid drug sales

```r
fit <- h02 |> model(auto = ARIMA(log(Cost)))
report(fit)
```

```
## Series: Cost
## Model: ARIMA(2,1,0)(0,1,1)[12]
## Transformation: log(Cost)
##
## Coefficients:
##           ar1      ar2     sma1
##       -0.8491  -0.4207  -0.6401
## s.e.   0.0712   0.0714   0.0694
##
## sigma^2 estimated as 0.004387:  log likelihood=245
## AIC=-483   AICc=-483   BIC=-470
```

Corticosteroid drug sales

```
gg_tsresiduals(fit)
```



Corticosteroid drug sales

```r
augment(fit) |>
  features(.innov, ljung_box, lag = 36, dof = 3)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 auto      59.3   0.00332
```

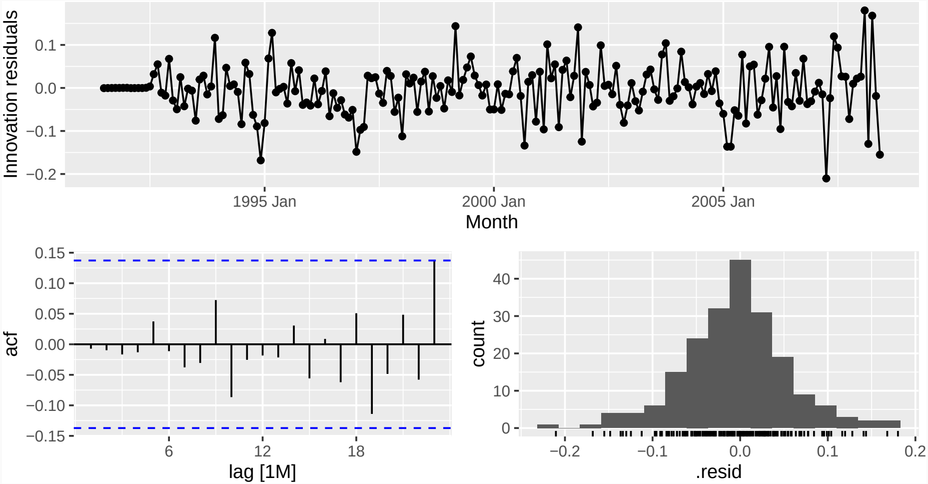
Corticosteroid drug sales

```r
fit <- h02 |>
  model(best = ARIMA(log(Cost),
    stepwise = FALSE, approximation = FALSE,
    order_constraint = p + q + P + Q <= 9))
report(fit)
```

```
## Series: Cost
## Model: ARIMA(4,1,1)(2,1,2)[12]
## Transformation: log(Cost)
##
## Coefficients:
##           ar1    ar2    ar3     ar4     ma1   sar1    sar2    sma1   sma2
##       -0.0425  0.210  0.202  -0.227  -0.742  0.621  -0.383  -1.202  0.496
## s.e.   0.2167  0.181  0.114   0.081   0.207  0.242   0.118   0.249  0.213
##
## sigma^2 estimated as 0.004049:  log likelihood=254
## AIC=-489   AICc=-487   BIC=-456
```

Corticosteroid drug sales

```
gg_tsresiduals(fit)
```



Corticosteroid drug sales

```r
augment(fit) |>
  features(.innov, ljung_box, lag = 36, dof = 9)
```

```
## # A tibble: 1 x 3
##   .model lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 best      36.5     0.106
```

Corticosteroid drug sales

Training data: July 1991 to June 2006

Test data: July 2006–June 2008

```
fit <- h02 |>
  filter_index(~ "2006 Jun") |>
  model(
    ARIMA(log(Cost) ~ 0 + pdq(3, 0, 0) + PDQ(2, 1, 0)),
    ARIMA(log(Cost) ~ 0 + pdq(3, 0, 1) + PDQ(2, 1, 0)),
    ARIMA(log(Cost) ~ 0 + pdq(3, 0, 2) + PDQ(2, 1, 0)),
    ARIMA(log(Cost) ~ 0 + pdq(3, 0, 1) + PDQ(1, 1, 0))
    # ... #
  )

fit |>
  forecast(h = "2 years") |>
  accuracy(h02)
```

Corticosteroid drug sales

.model	RMSE
ARIMA(3,0,1)(1,1,1)[12]	0.0619
ARIMA(3,0,1)(0,1,2)[12]	0.0621
ARIMA(3,0,1)(0,1,1)[12]	0.0630
ARIMA(2,1,0)(0,1,1)[12]	0.0630
ARIMA(4,1,1)(2,1,2)[12]	0.0631
ARIMA(3,0,2)(2,1,0)[12]	0.0651
ARIMA(3,0,1)(2,1,0)[12]	0.0653
ARIMA(3,0,1)(1,1,0)[12]	0.0666
ARIMA(3,0,0)(2,1,0)[12]	0.0668

Corticosteroid drug sales

- Models with lowest AICc values tend to give slightly better results than the other models.
- AICc comparisons must have the same orders of differencing. But RMSE test set comparisons can involve any models.
- Use the best model available, even if it does not pass all tests.

Corticosteroid drug sales

```
fit <- h02 |>
  model(ARIMA(Cost ~ 0 + pdq(3, 0, 1) + PDQ(0, 1, 2)))
fit |>
  forecast() |>
  autoplot(h02) + labs(y = "H02 Expenditure ($AUD)")
```