Seasonal Arima

Lecture 11

Dr. Colin Rundel

ARIMA - General Guidance

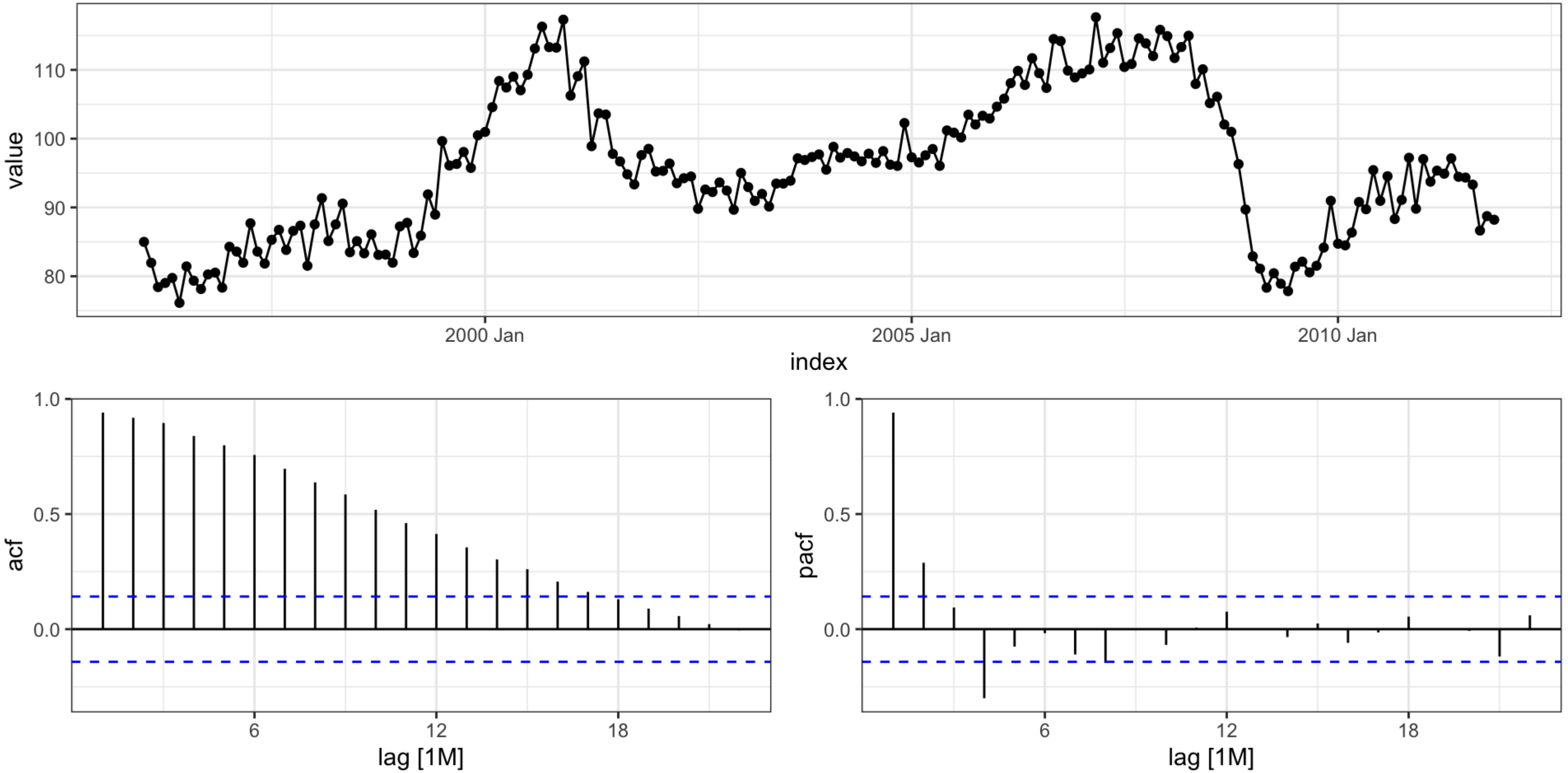
1. Positive autocorrelations out to a large number of lags usually indicates a need for differencing
2. Slightly too much or slightly too little differencing can be corrected by adding AR or MA terms respectively.
3. A model with no differencing usually includes a constant term, a model with two or more orders (rare) differencing usually does not include a constant term.
4. After differencing, if the PACF has a sharp cutoff then consider adding AR terms to the model.
5. After differencing, if the ACF has a sharp cutoff then consider adding an MA term to the model.
6. It is possible for an AR term and an MA term to cancel each other's effects, so try models with fewer AR terms and fewer MA terms.

Based on guidance in Robert Nau's notes on regression and time series analysis - Chapter 5

Electrical Equipment Sales

Data

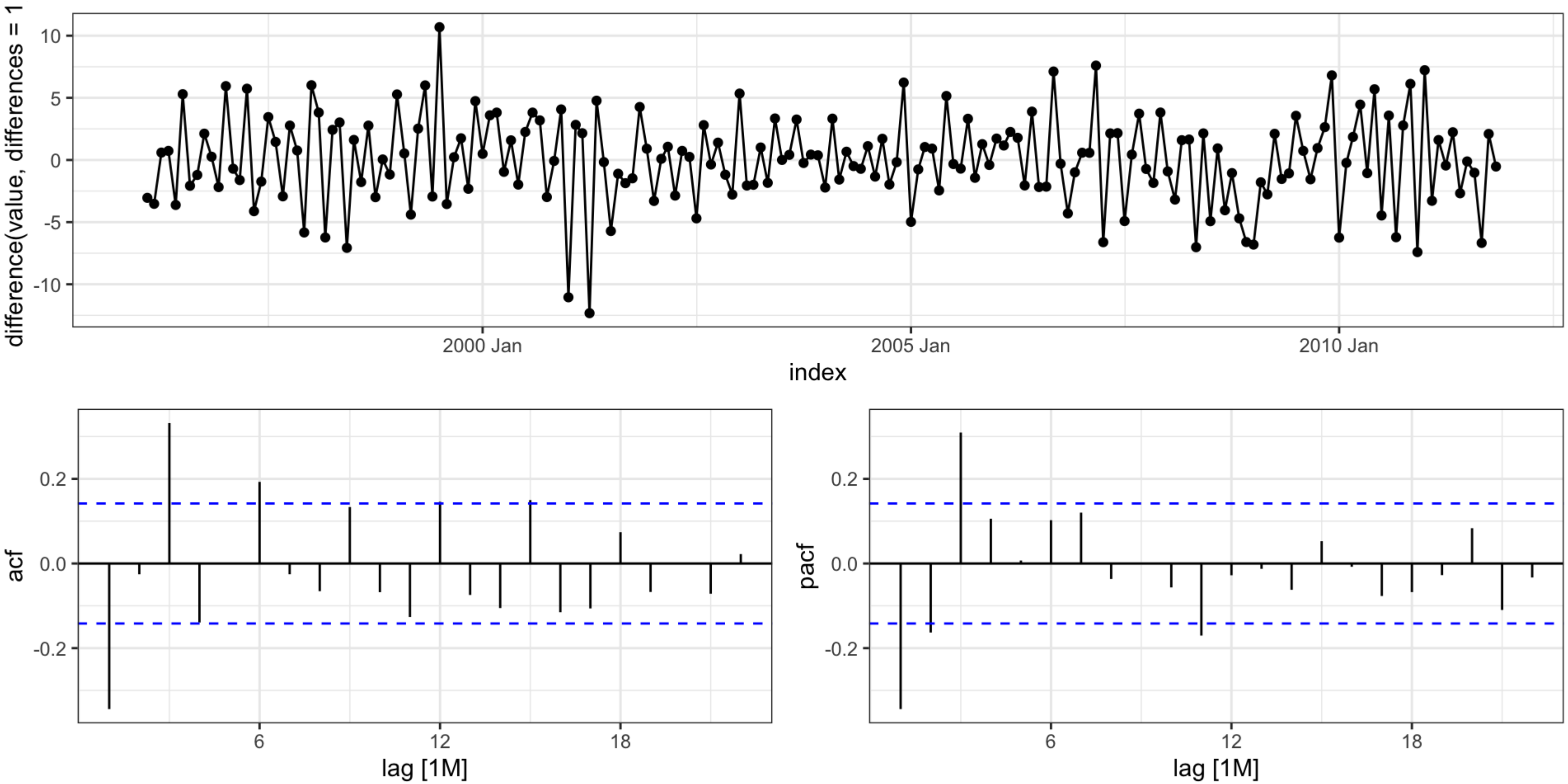
```r
feasts::gg_tsdisplay(elec_sales, y=value, plot_type="partial")
```



Differencing

`differences = 1` `differences = 2`

```r
feasts::gg_tsdisplay(elec_sales, y=difference(value, differences = 1), plot_type="partial")
```



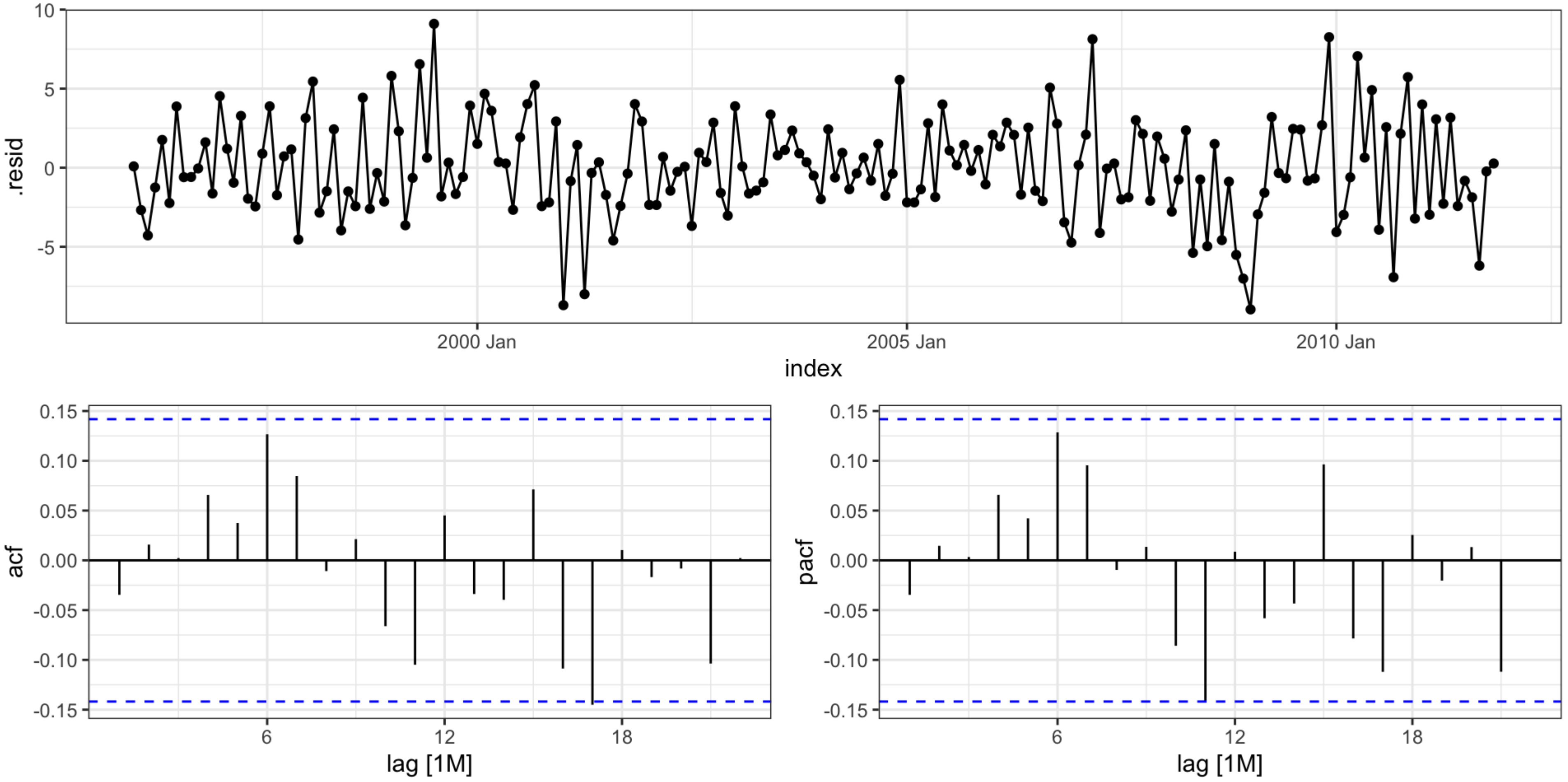
Model

```r
m = model(
  elec_sales,
  ARIMA(value ~ pdq(3,1,0))
)

glance(m)
```

```
# A tibble: 1 × 8
  .model                          sigma2 log_lik   AIC  AICc   BIC ar_roots
ma_roots
  <chr>                            <dbl>   <dbl> <dbl> <dbl> <dbl> <list>
<list>
1 ARIMA(value ~ pdq(3, 1, 0))   9.85    -486.  979.  980.  992. <cpl>
<cpl>
```

Residuals

```
residuals(m) |>
  feasts::gg_tsdisplay(y=.resid, plot_type="partial")
```



Information Criteria

The fable package provides a number of different information criteria for model selection,

$$\text{AIC} = -2 \log \mathcal{L} + 2(p + q + k + 1)$$

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{n - p - q - k - 2}$$

$$\text{BIC} = \text{AIC} + (\log(n) - 2)(p + q + k + 1)$$

For small values of n, AIC can overfit the data (select too many predictors) and so AICc is preferred. BIC is a more conservative version of AIC that penalizes the number of predictors more heavily.

Note that since differencing changes the data we are fitting the ARMA model to, it does not make sense to

Model Comparison

Model choices:

```
model(
  elec_sales,
  ARIMA(value ~ pdq(3,1,0)),
  ARIMA(value ~ pdq(2,1,0)),
  ARIMA(value ~ pdq(4,1,0)),
  ARIMA(value ~ pdq(3,1,1))
) |>
  glance()
```

```
# A tibble: 4 × 8
  .model                          sigma2 log_lik   AIC  AICc   BIC ar_roots ma_roots
  <chr>                            <dbl>   <dbl> <dbl> <dbl> <dbl> <list>   <list>
1 ARIMA(value ~ pdq(3, 1, 0))   9.85    -486.  979.  980.  992. <cpl>    <cpl>
2 ARIMA(value ~ pdq(2, 1, 0))  10.9     -495.  997.  997. 1006. <cpl>    <cpl>
3 ARIMA(value ~ pdq(4, 1, 0))   9.78    -484.  979.  979.  995. <cpl>    <cpl>
4 ARIMA(value ~ pdq(3, 1, 1))   9.74    -484.  978.  978.  994. <cpl>    <cpl>
```

Automatic selection (AICc)

```r
model(
  elec_sales,
  ARIMA(value)
) |>
  report()
```

```
Series: value
Model: ARIMA(3,1,1)

Coefficients:
         ar1     ar2     ar3      ma1
      0.0519  0.1191  0.3730  -0.4542
s.e.  0.1840  0.0888  0.0679   0.1993

sigma^2 estimated as 9.737:  log likelihood=-484.08
AIC=978.17   AICc=978.49   BIC=994.4
```

Automatic selection (AIC)

```r
model(
  elec_sales,
  ARIMA(value, ic = "aic")
) |>
  report()
```

```
Series: value
Model: ARIMA(3,1,1)

Coefficients:
         ar1     ar2     ar3      ma1
      0.0519  0.1191  0.3730  -0.4542
s.e.  0.1840  0.0888  0.0679   0.1993

sigma^2 estimated as 9.737:  log likelihood=-484.08
AIC=978.17   AICc=978.49   BIC=994.4
```

Automatic selection (BIC)

```r
model(
  elec_sales,
  ARIMA(value, ic = "bic")
) |>
  report()
```

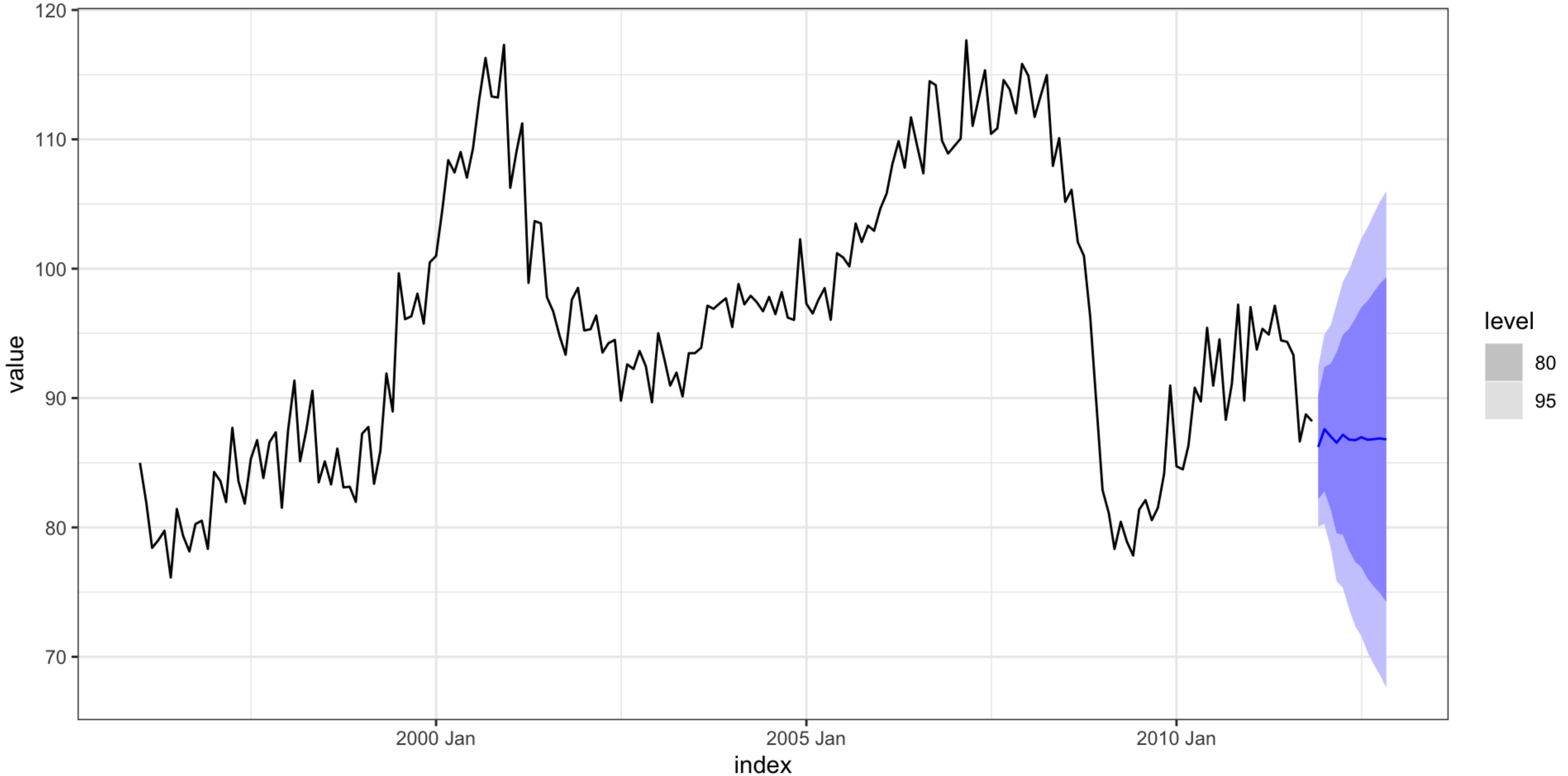
```
Series: value
Model: ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.7738  -1.2298  0.5106
s.e.  0.0933   0.1035  0.0695

sigma^2 estimated as 10.2:  log likelihood=-488.99
AIC=985.97   AICc=986.19   BIC=998.96
```

Model fit

```
m |>
  augment() |>
  ggplot(aes(x=as.Date(index))) +
    geom_line(aes(y=value), color = "black") +
    geom_line(aes(y=.fitted), color = "red", alpha=0.75)
```

Model forecast

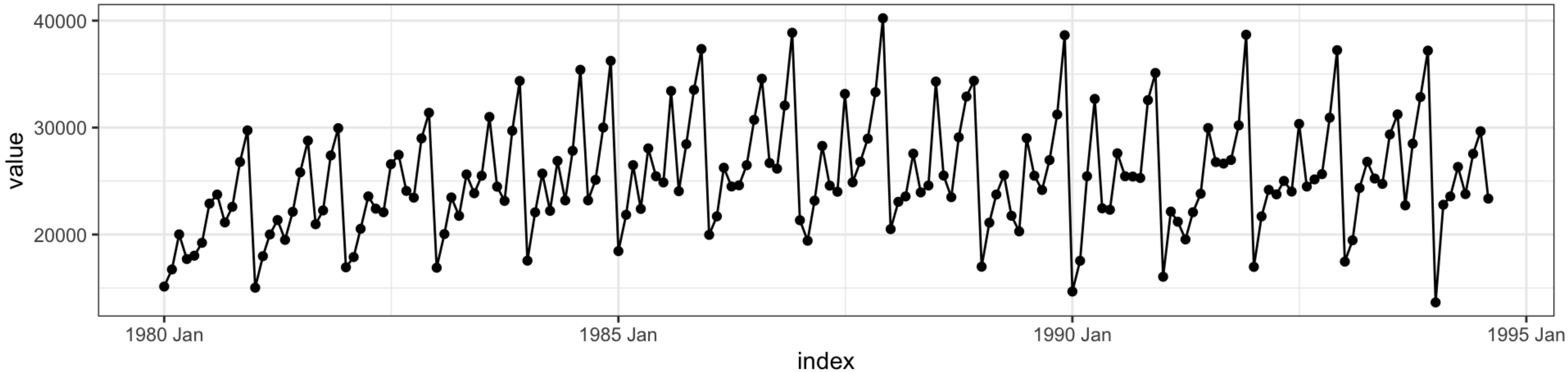
```
m |>
  forecast(h=12) |>
  autoplot(elec_sales)
```

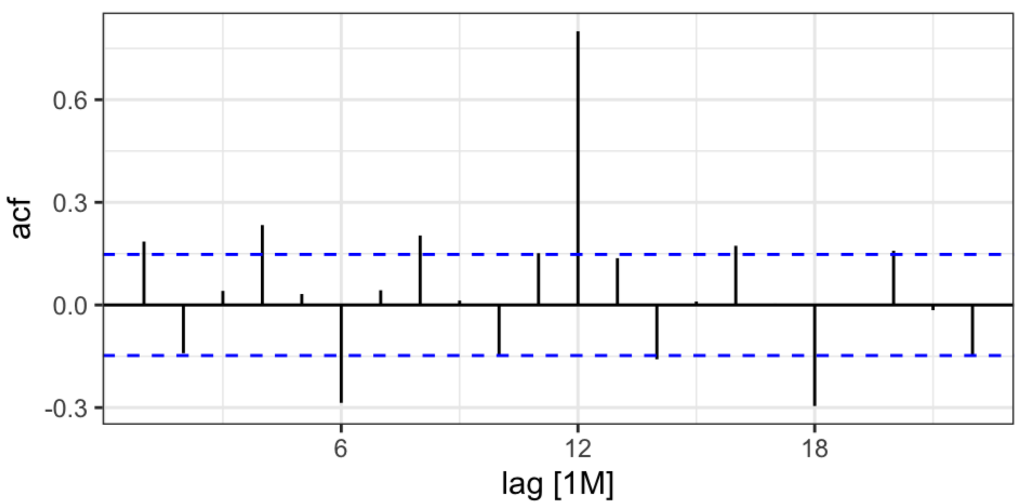


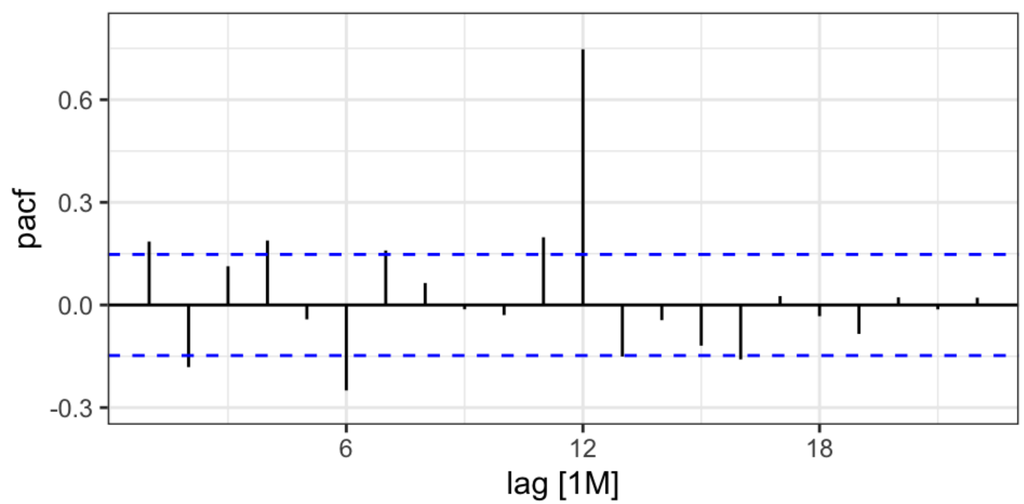
Seasonal Models

Australian Wine Sales Example

Australian total wine sales by wine makers in bottles <= 1 litre. Jan 1980 – Aug 1994.

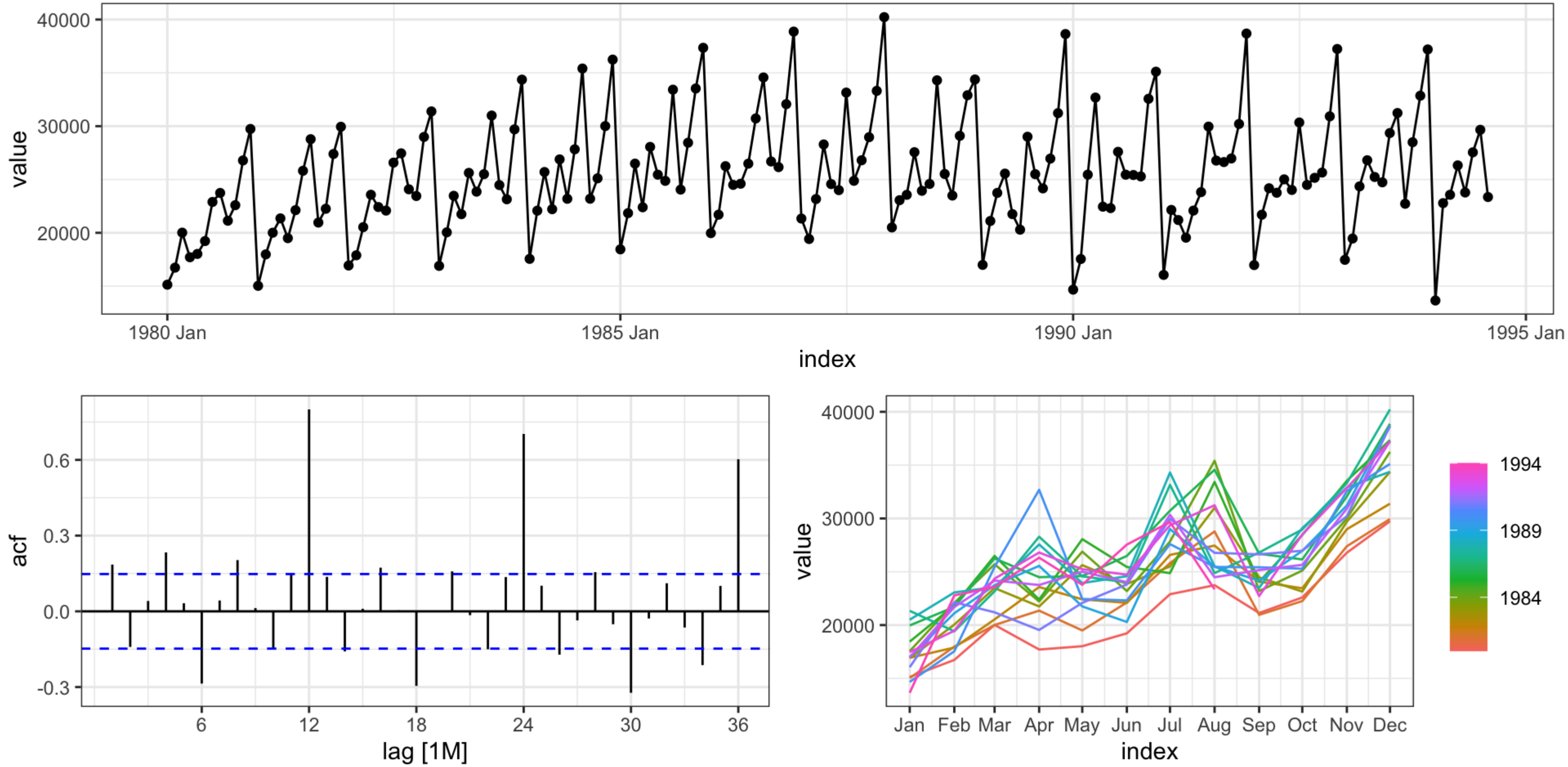






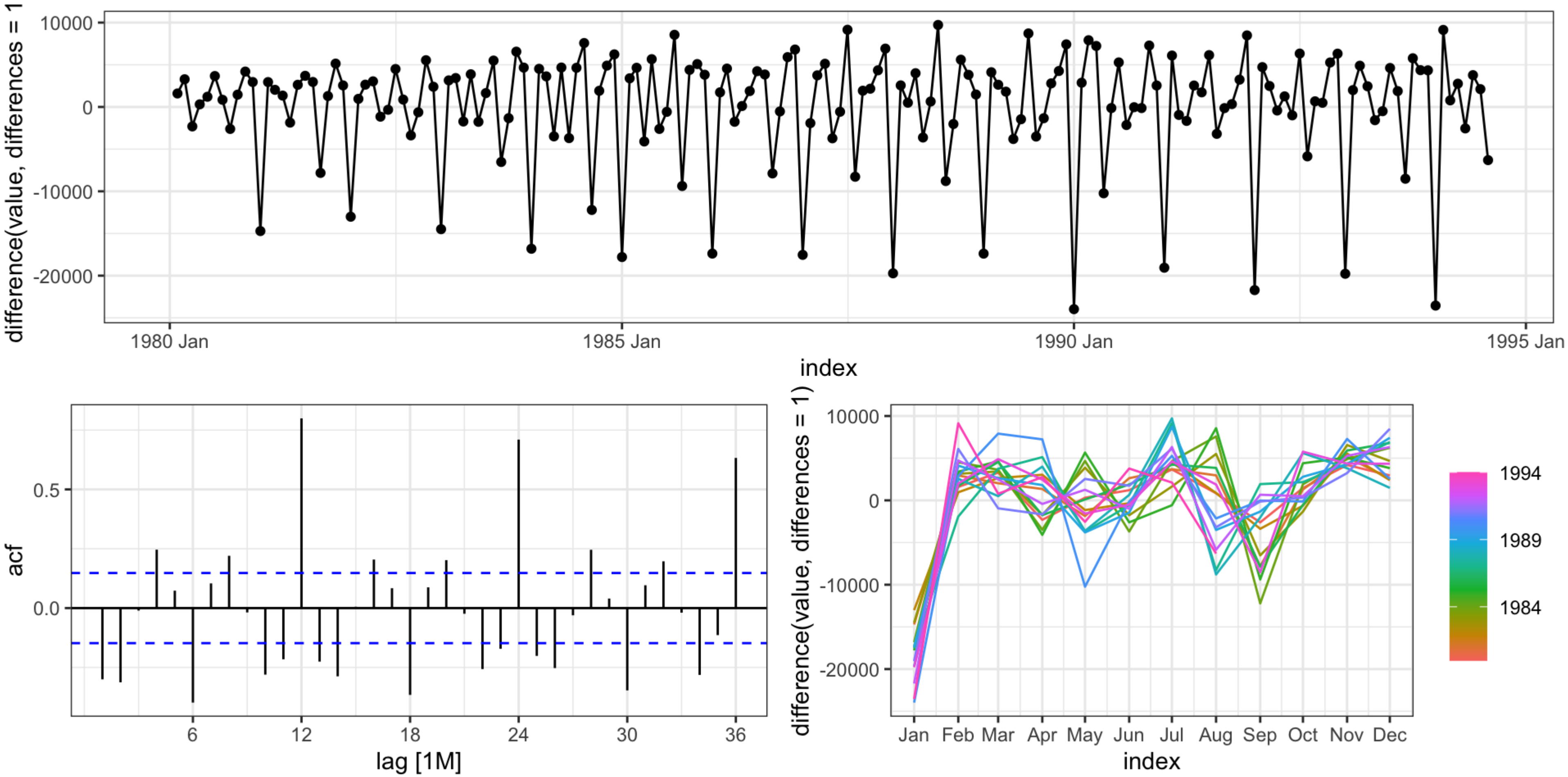
Seasonal plot

```r
feasts::gg_tsdisplay(wineind, y=value, lag_max = 36)
```



Differencing

```r
feasts::gg_tsdisplay(wineind, y=difference(value, differences = 1), lag_max = 36)
```



Seasonal ARIMA

We can extend the existing ARIMA model to handle these higher order lags (without including all of the intervening lags).

Seasonal $\text{ARIMA}\,(p, d, q) \times (P, D, Q)_s$:

$$\Phi_P(L^s)\,\phi_p(L)\,\Delta_s^D\,\Delta^d\,y_t = \delta + \Theta_Q(L^s)\,\theta_q(L)\,w_t$$

where

$$\begin{aligned}
\phi_p(L) &= 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \\
\theta_q(L) &= 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^q \\
\Delta^d &= (1 - L)^d
\end{aligned}$$

$$\begin{aligned}
\Phi_P(L^s) &= 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \ldots - \Phi_P L^{Ps} \\
\Theta_Q(L^s) &= 1 + \Theta_1 L + \Theta_2 L^{2s} + \ldots + \theta_p L^{Qs} \\
\Delta_s^D &= (1 - L^s)^D
\end{aligned}$$

Seasonal Arima - Diff

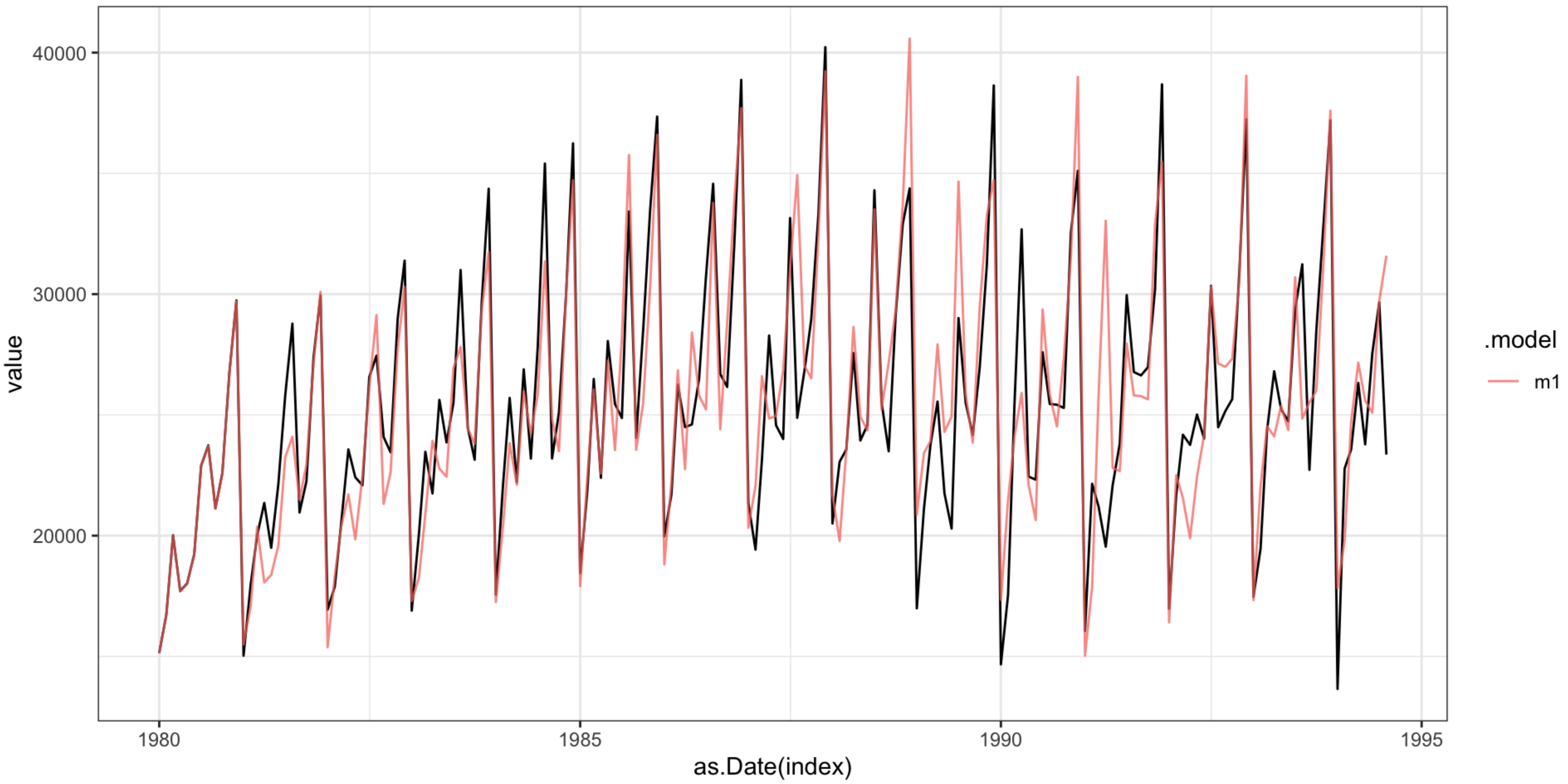
Lets consider an $\text{ARIMA}(0,0,0) \times (0,1,0)_{12}$:

$$
\begin{aligned}
(1 - L^{12})\, y_t &= \delta + w_t \\
y_t &= y_{t-12} + \delta + w_t
\end{aligned}
$$

```r
m = model(
  wineind,
  m1 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,0, period=12))
)
report(m)
```

```
Series: value
Model: ARIMA(0,0,0)(0,1,0)[12] w/ drift

Coefficients:
      constant
      355.0122
s.e.  208.5520

sigma^2 estimated as 7176802:  log likelihood=-1526.69
AIC=3057.37   AICc=3057.45   BIC=3063.57
```

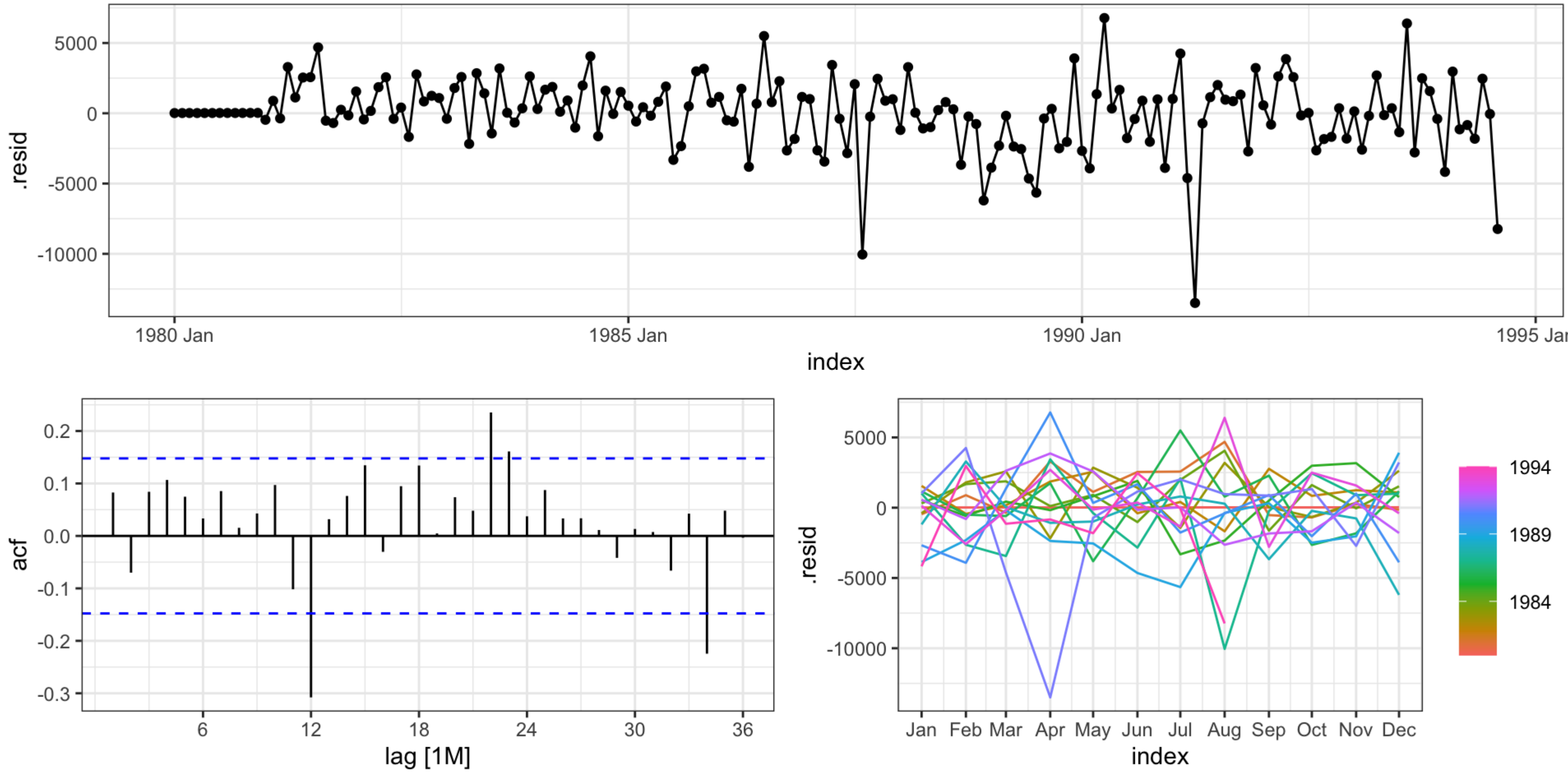
Fitted - Model 1



Residuals

Seasonal plot | pACF plot

```r
residuals(m) |>
  feasts::gg_tsdisplay(y=.resid, lag_max=36)
```



Model 2

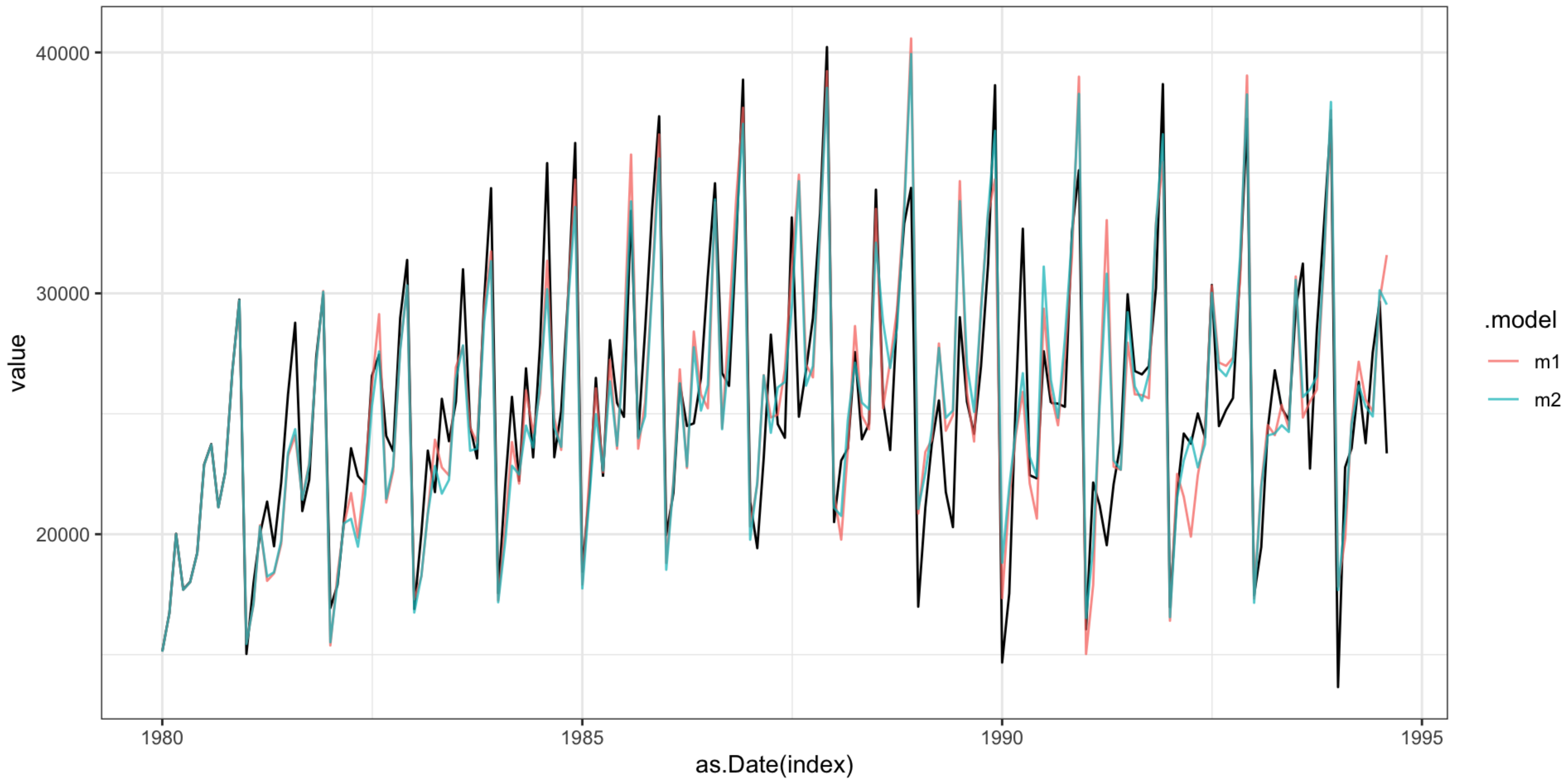
ARIMA$(0, 0, 0) \times (0, 1, 1)_{12}$:

$$
(1 - L^{12})y_t = \delta + (1 + \Theta_1 L^{12})w_t
$$
$$
y_t = \delta + y_{t-12} + w_t + \Theta_1 w_{t-12}
$$

```
m = model(
  wineind,
  m1 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,0, period=12)),
  m2 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,1, period=12)),
)
glance(m)
```

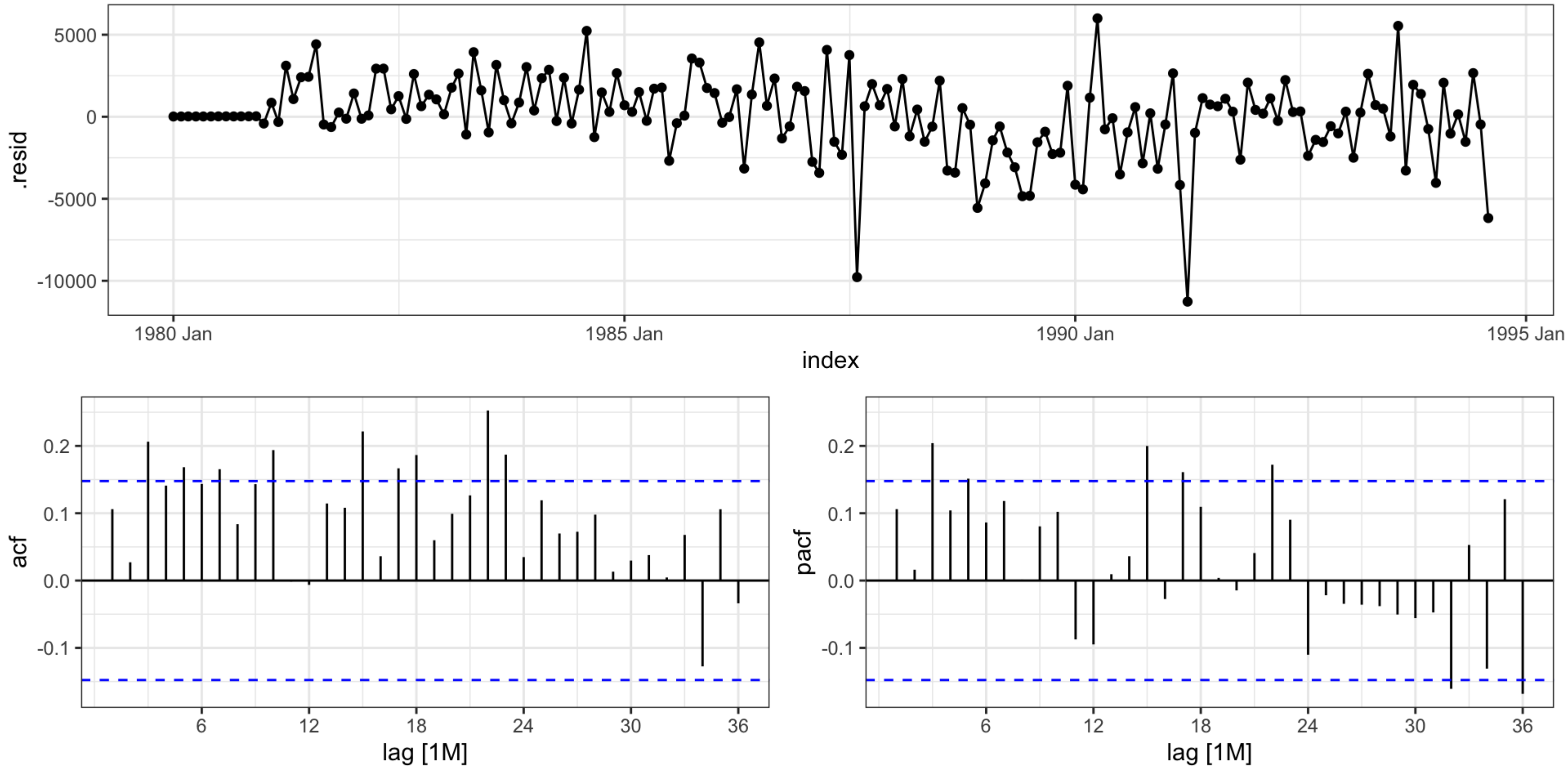
```
# A tibble: 2 × 8
  .model   sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
  <chr>     <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
1 m1     7176802.  -1527. 3057. 3057. 3064. <cpl [0]> <cpl [0]>
2 m2     6369103.  -1517. 3041. 3041. 3050. <cpl [0]> <cpl [12]>
```

Fitted - Model 2



Residuals

```r
residuals(m) |>
  filter(.model == "m2") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```



Model 3

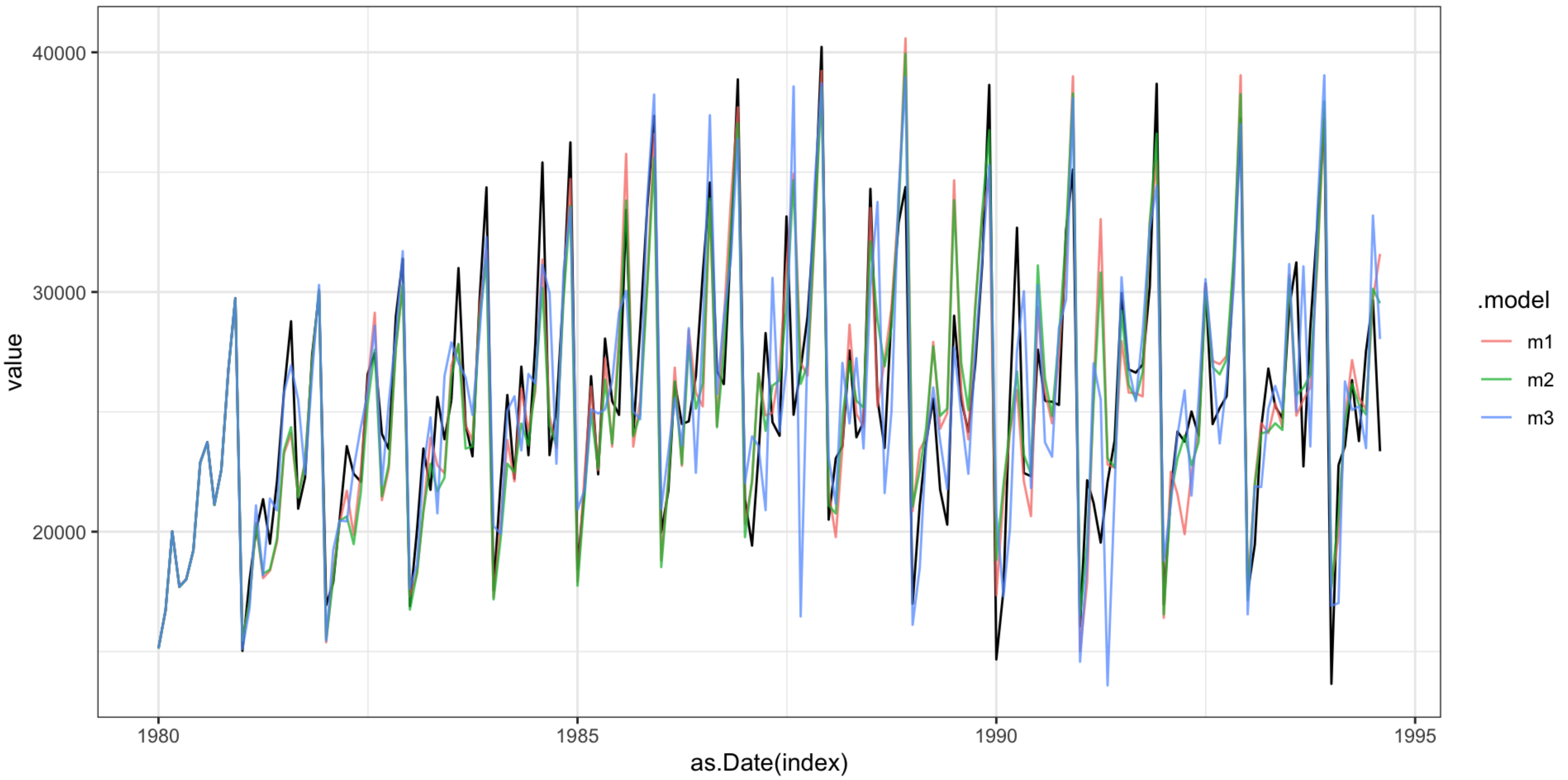
$\text{ARIMA}(0,1,0) \times (0,1,1)_{12}$

$$(1-L)\,(1-L^{12})y_t = \delta + (1+\Theta_1 L)w_t$$
$$y_t = \delta + y_{t-1} + y_{t-12} - y_{t-13} + w_t + w_{t-12}$$

```
m = model(
  wineind,
  m1 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,0, period=12)),
  m2 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,1, period=12)),
  m3 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,1, period=12))
)
glance(m)
```

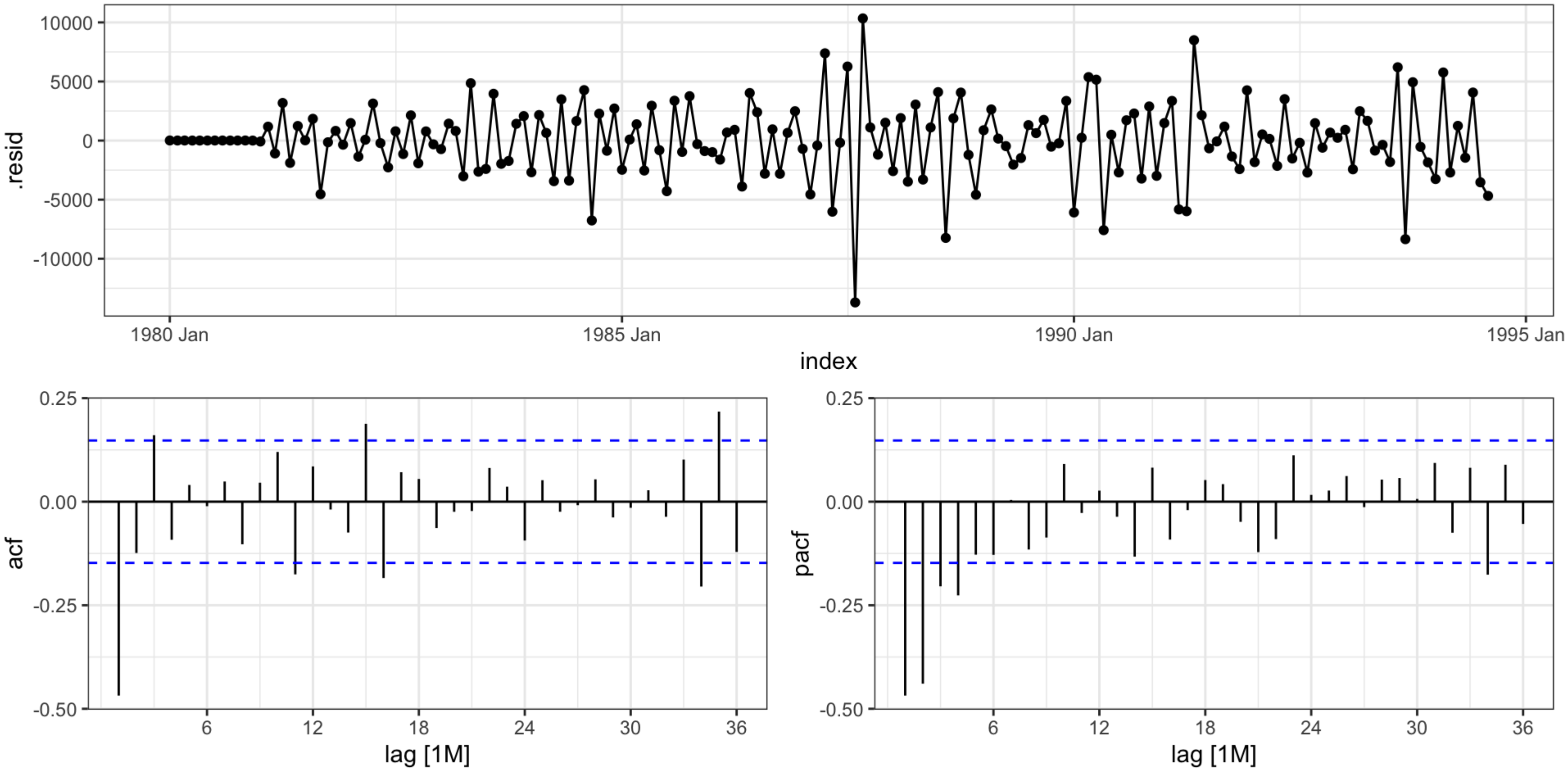
```
# A tibble: 3 × 8
  .model    sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
  <chr>      <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
1 m1      7176802.  -1527. 3057. 3057. 3064. <cpl [0]> <cpl [0]>
2 m2      6369103.  -1517. 3041. 3041. 3050. <cpl [0]> <cpl [12]>
3 m3     10751355.  -1553. 3109. 3109. 3115. <cpl [0]> <cpl [12]>
```

Fitted model



Residuals

```
residuals(m) |>
  filter(.model == "m3") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```



Model 4

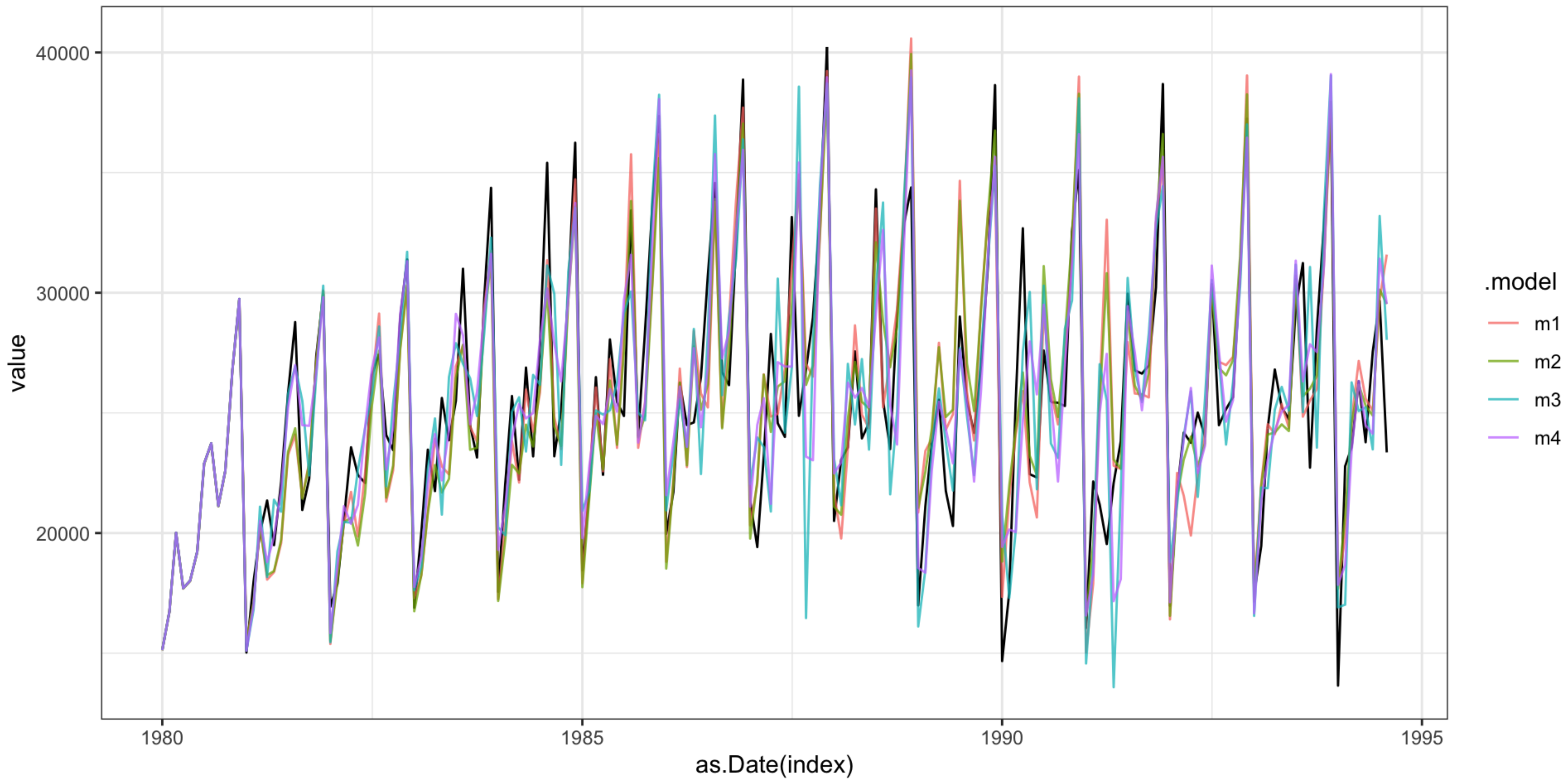
$\text{ARIMA}(1,1,0) \times (0,1,1)_{12}$

$$(1-\phi_1 L)\,(1-L)\,(1-L^{12})y_t = \delta + (1+\Theta_1 L)w_t$$
$$y_t = \delta + (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + y_{t-12} - (1+\phi_1)y_{t-13} + \phi_1 y_{t-14} + w_t + w_{t-12}$$

```r
m = model(
  wineind,
  m1 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,0, period=12)),
  m2 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,1, period=12)),
  m3 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,1, period=12)),
  m4 = ARIMA(value ~ pdq(1,1,0) + PDQ(0,1,1, period=12))
)
glance(m)
```

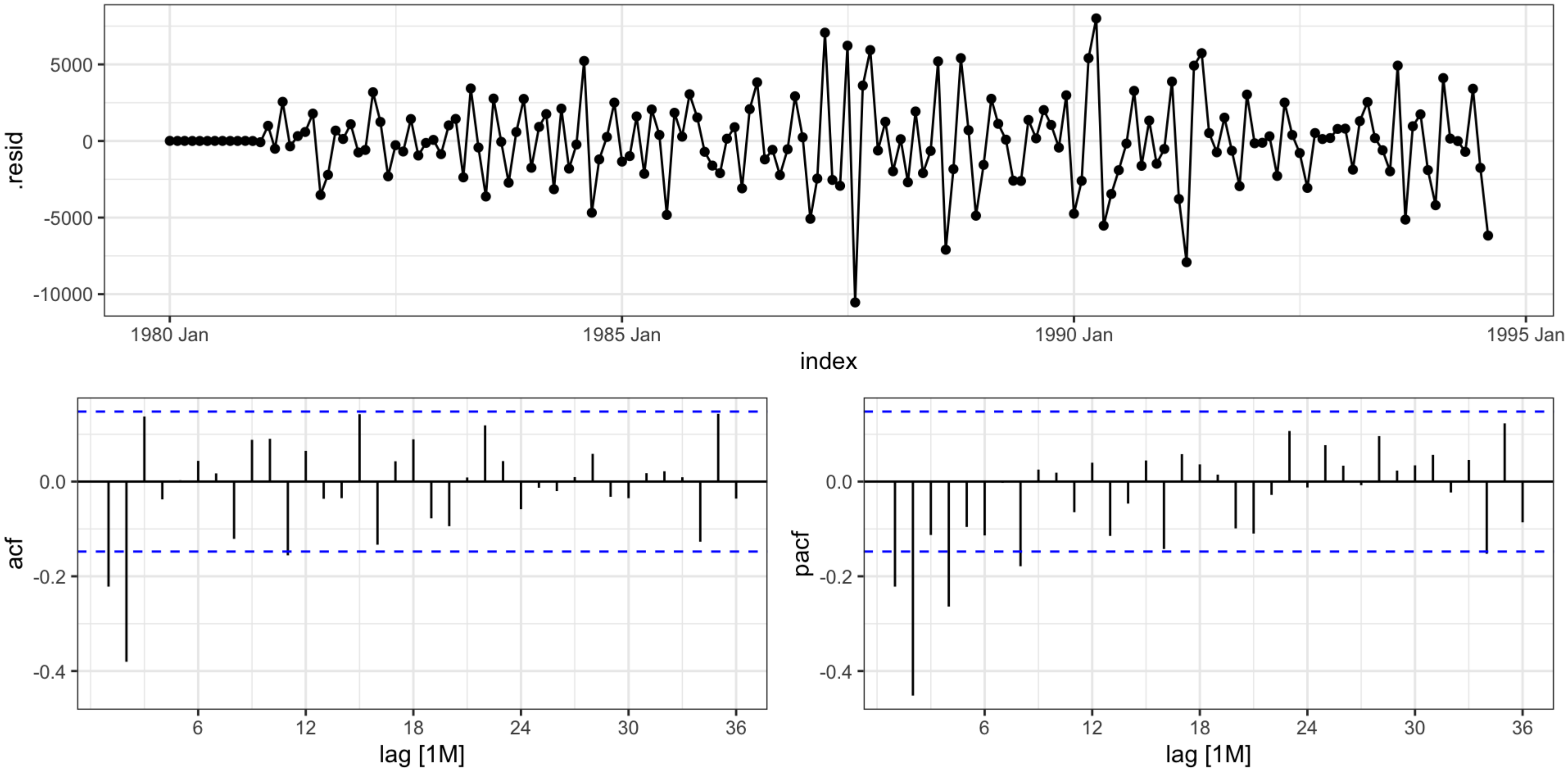
```
# A tibble: 4 × 8
  .model     sigma2 log_lik   AIC  AICc   BIC ar_roots  ma_roots
  <chr>       <dbl>   <dbl> <dbl> <dbl> <dbl> <list>    <list>
1 m1       7176802.  -1527. 3057. 3057. 3064. <cpl [0]> <cpl [0]>
2 m2       6369103.  -1517. 3041. 3041. 3050. <cpl [0]> <cpl [12]>
3 m3      10751355.  -1553. 3109. 3109. 3115. <cpl [0]> <cpl [12]>
4 m4       8267920.  -1532. 3070. 3070. 3079. <cpl [1]> <cpl [12]>
```

Fitted model



Residuals

```
residuals(m) |>
  filter(.model == "m4") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```



Model 5

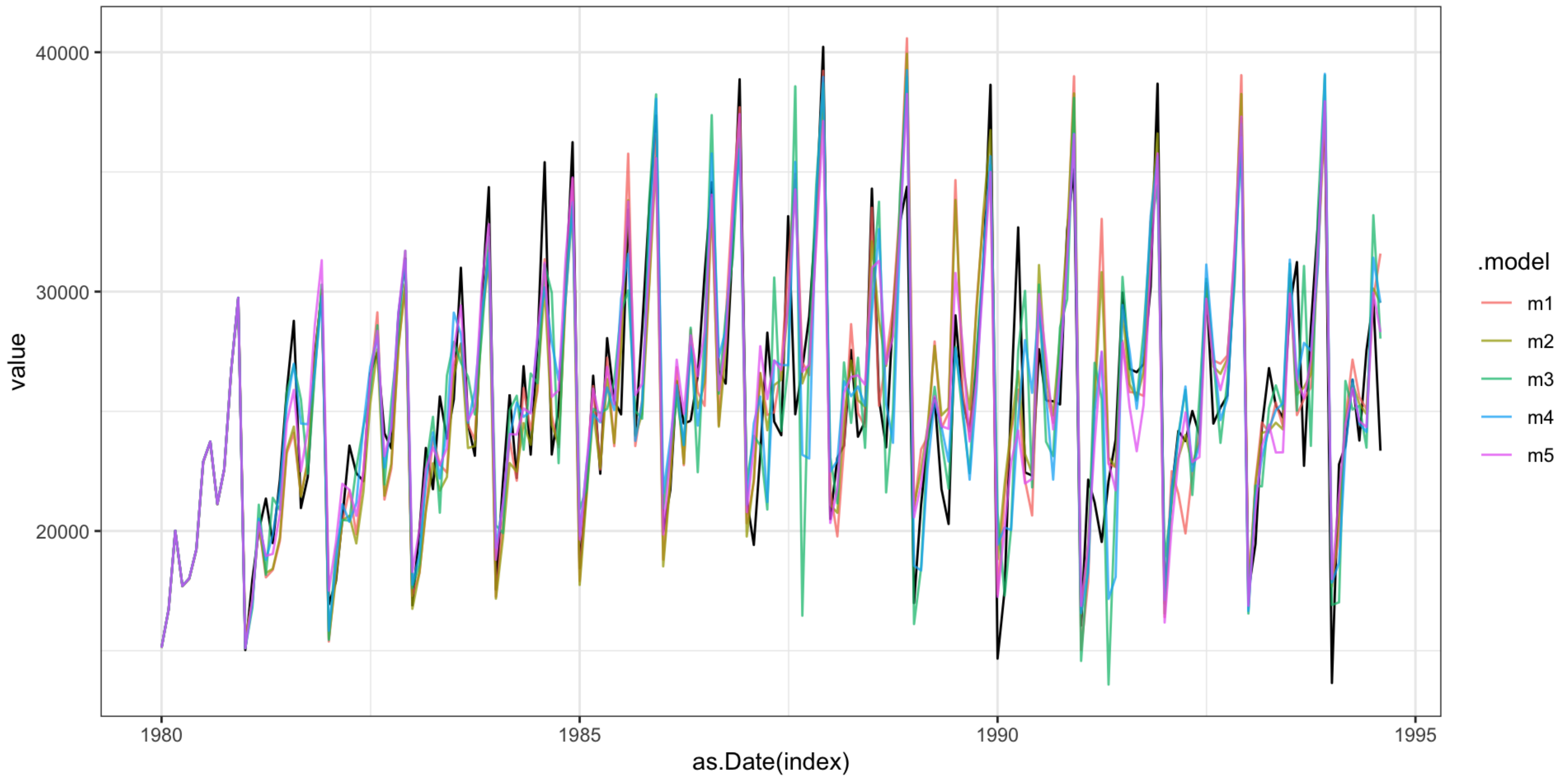
ARIMA$(1,1,2) \times (0,1,1)_{12}$

$$
\begin{aligned}
(1-\phi_1 L)\,(1-L)\,(1-L^{12})y_t &= \delta + (1+\theta_1 L + \theta_2 L^2)(1+\Theta_1 L)w_t \\
y_t = \delta + (1+\phi_1)y_{t-1} &- \phi_1 y_{t-2} + y_{t-12} - (1+\phi_1)y_{t-13} + \phi_1 y_{t-14} \\
&+ w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + w_{t-12} + \theta_1 w_{t-13} + \theta_2 w_{t-14}
\end{aligned}
$$

```r
m = model(wineind,
  m1 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,0, period=12)),
  m2 = ARIMA(value ~ pdq(0,0,0) + PDQ(0,1,1, period=12)),
  m3 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,1, period=12)),
  m4 = ARIMA(value ~ pdq(1,1,0) + PDQ(0,1,1, period=12)),
  m5 = ARIMA(value ~ pdq(1,1,2) + PDQ(0,1,1, period=12))
); glance(m)
```

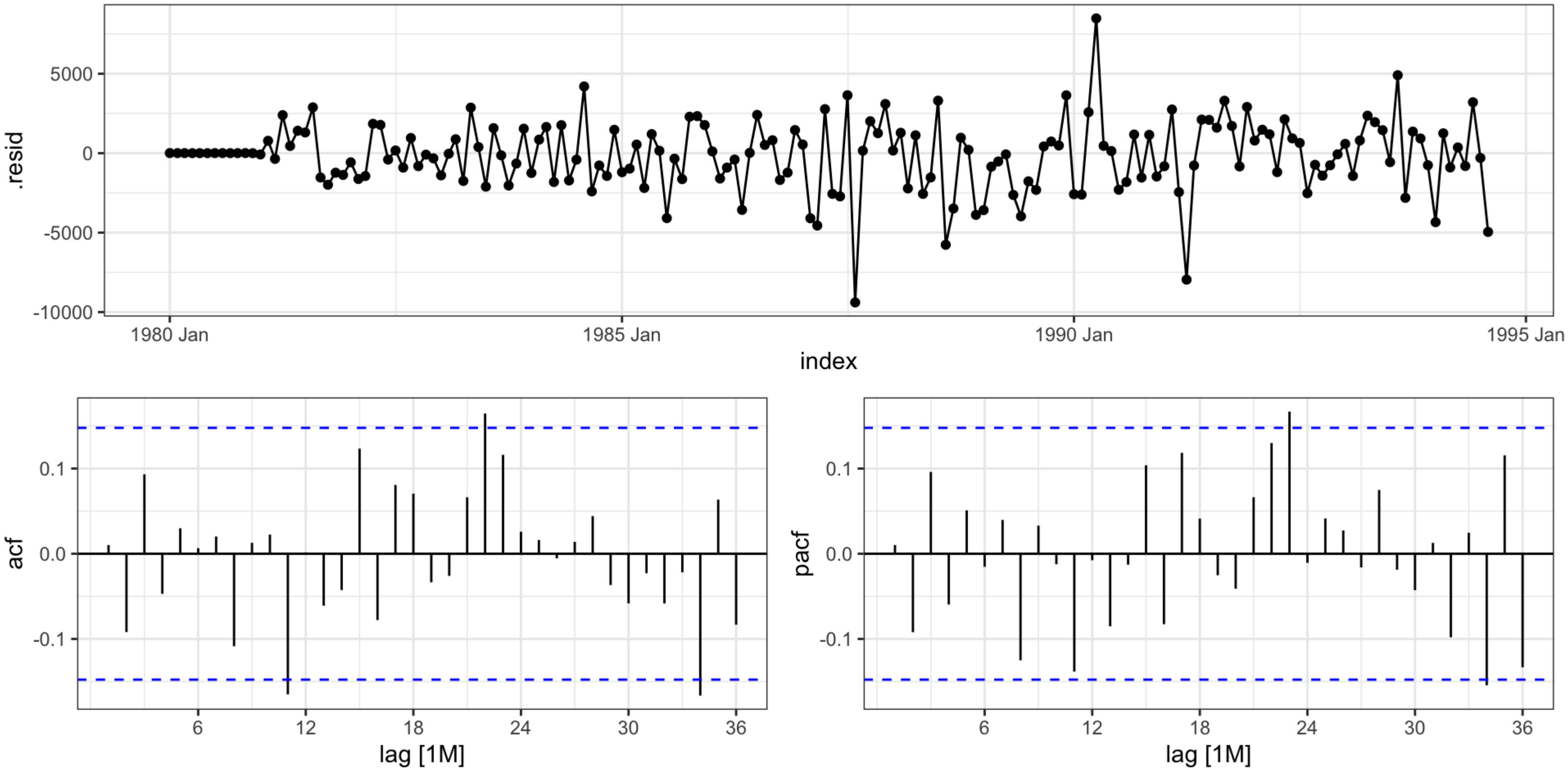
```
# A tibble: 5 × 8
  .model     sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
  <chr>       <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
1 m1       7176802.  -1527. 3057. 3057. 3064. <cpl [0]> <cpl [0]>
2 m2       6369103.  -1517. 3041. 3041. 3050. <cpl [0]> <cpl [12]>
3 m3      10751355.  -1553. 3109. 3109. 3115. <cpl [0]> <cpl [12]>
4 m4       8267920.  -1532. 3070. 3070. 3079. <cpl [1]> <cpl [12]>
5 m5       5399312.  -1497. 3004. 3004. 3020. <cpl [1]> <cpl [14]>
```

Fitted model



Residuals

```
residuals(m) |>
  filter(.model == "m5") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```



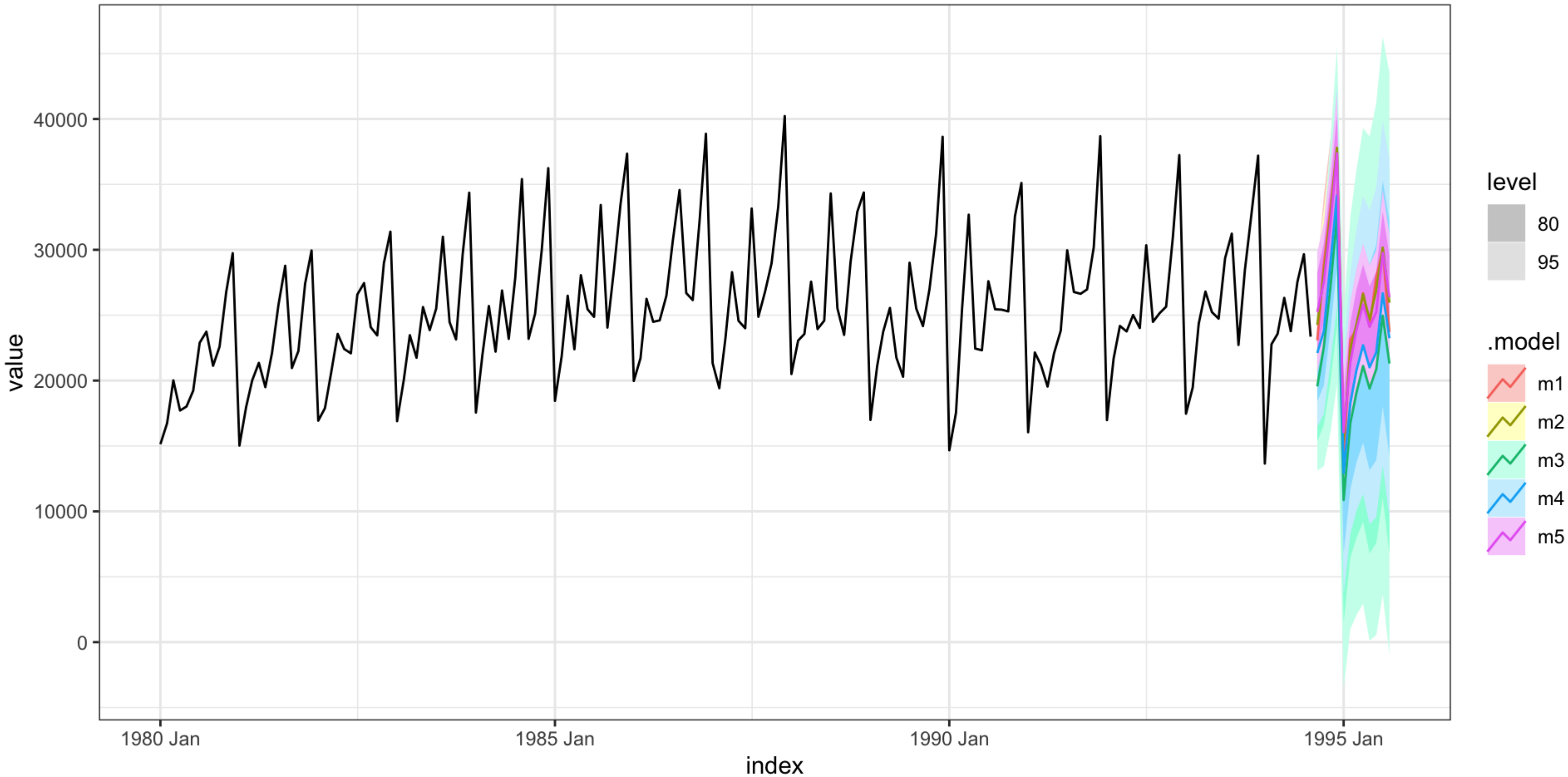
Automated selection

```r
m_auto = model(
  wineind,
  ARIMA(value)
)

report(m_auto)
```

```
Series: value
Model: ARIMA(1,1,2)(0,1,1)[12]

Coefficients:
         ar1      ma1     ma2     sma1
      0.4299  -1.4673  0.5339  -0.6600
s.e.  0.2984   0.2658  0.2340   0.0799

sigma^2 estimated as 5399312:  log likelihood=-1497.05
AIC=3004.1   AICc=3004.48   BIC=3019.57
```

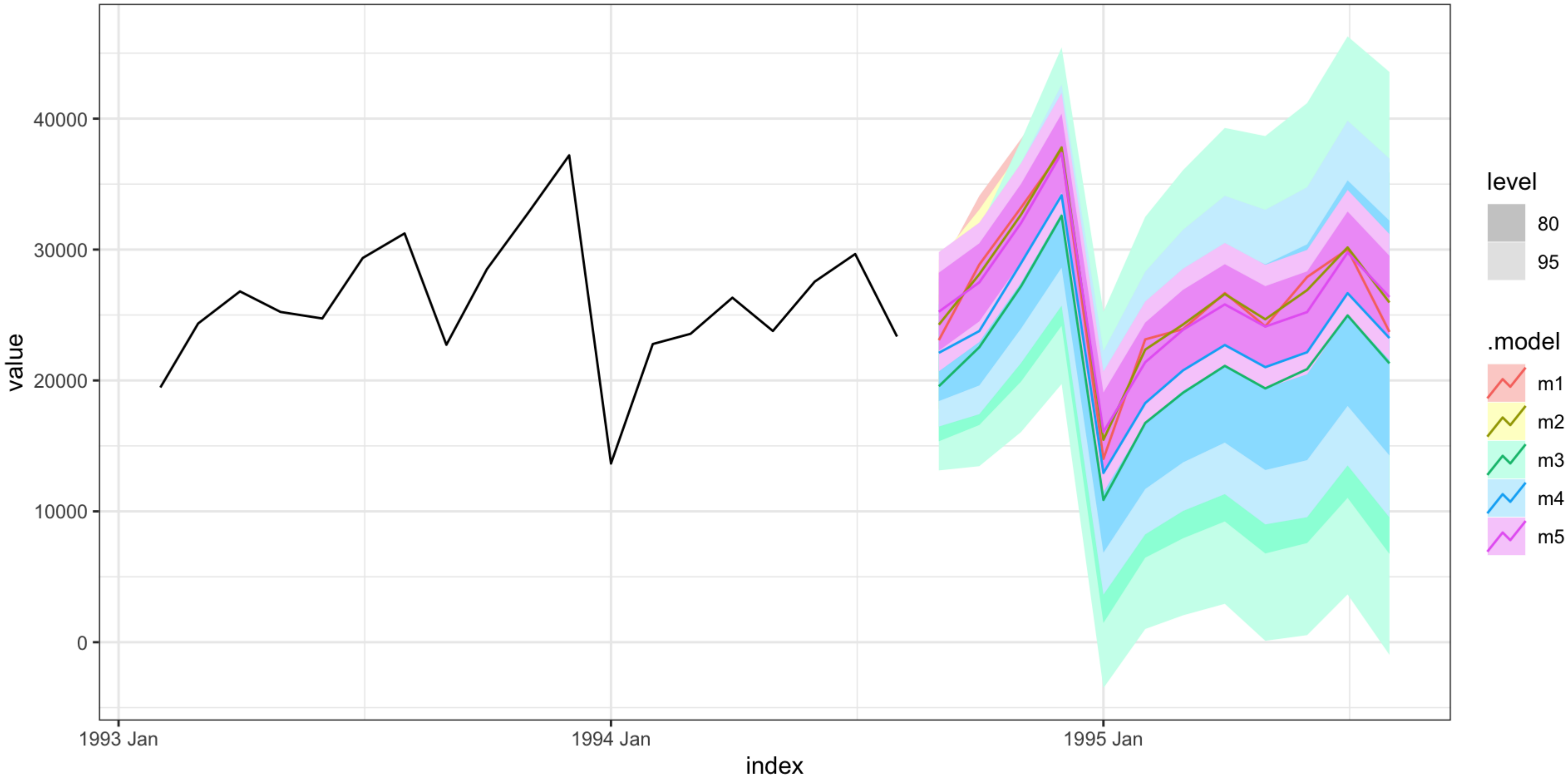
Forecasts

```r
m |>
  forecast(h=12) |>
  autoplot(wineind)
```



Forecasts (1993-1996)

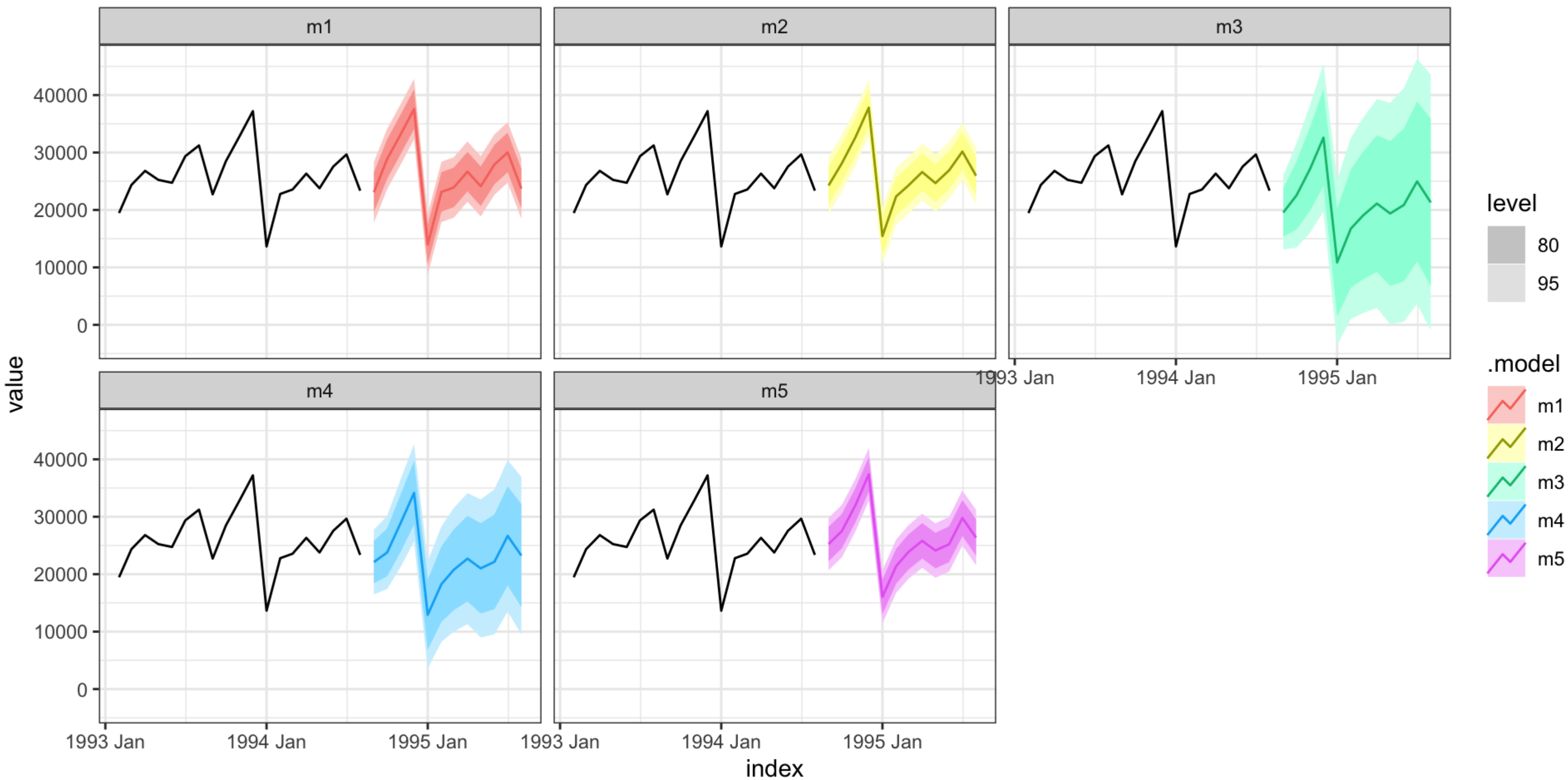
```
m |>
  forecast(h=12) |>
  autoplot(
    wineind |>
      filter(index > make_yearmonth(1993, 1))
  )
```



value
40000
30000
20000
10000
0
1993 Jan
1994 Jan
1995 Jan
index
level
80
95
.model
m1
m2
m3
m4
m5

Forecasts

```r
m |>
  forecast(h=12) |>
  autoplot(wineind |> filter(index > make_yearmonth(1993, 1))) +
    facet_wrap(~.model)
```

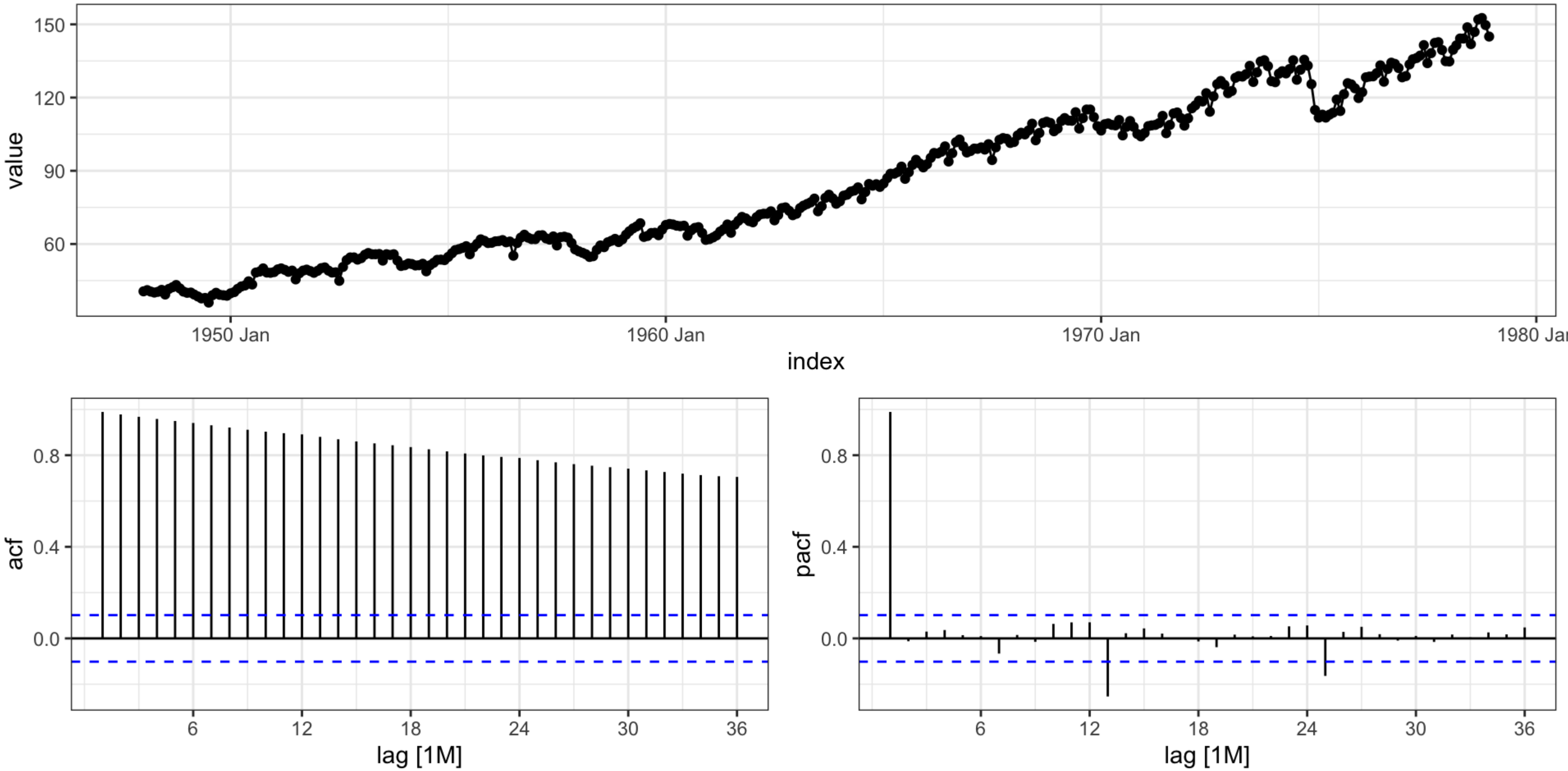


Federal Reserve Board Production Index

`prodn` from the `astsa` package

Monthly Federal Reserve Board Production Index (1948-1978)

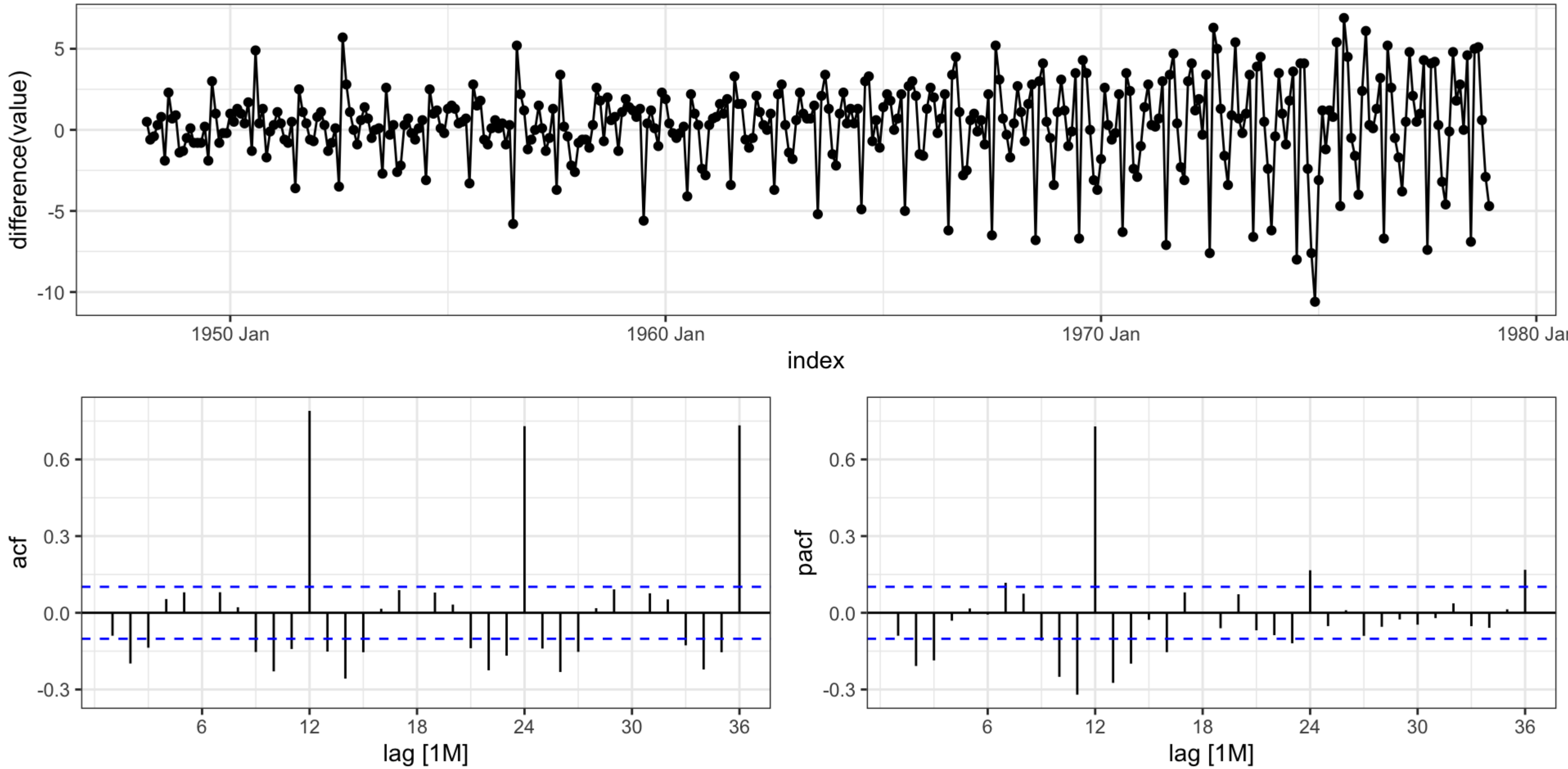
```
feasts::gg_tsdisplay(prodn, y=value, plot_type = "partial", lag_max=36)
```



Differencing

Based on the ACF it seems like standard differencing may be required

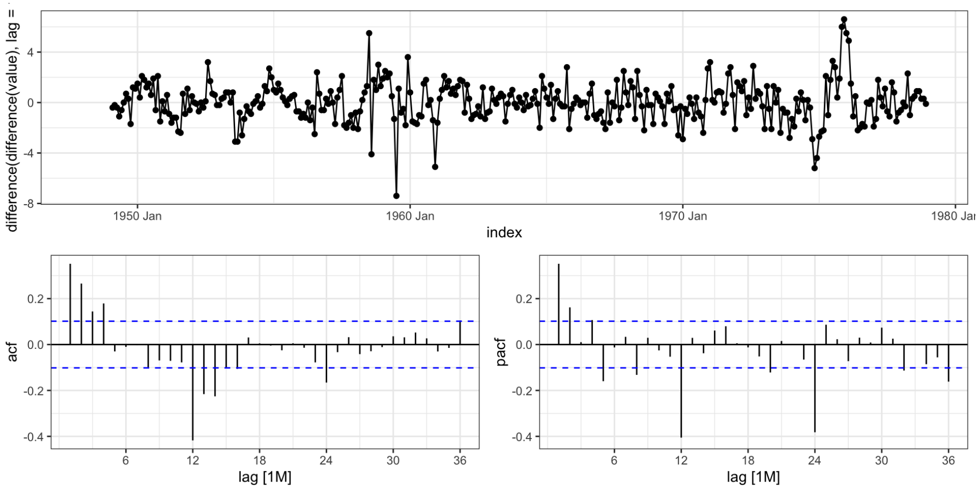
```
feasts::gg_tsdisplay(prodn, y=difference(value), plot_type = "partial", lag_max=36)
```

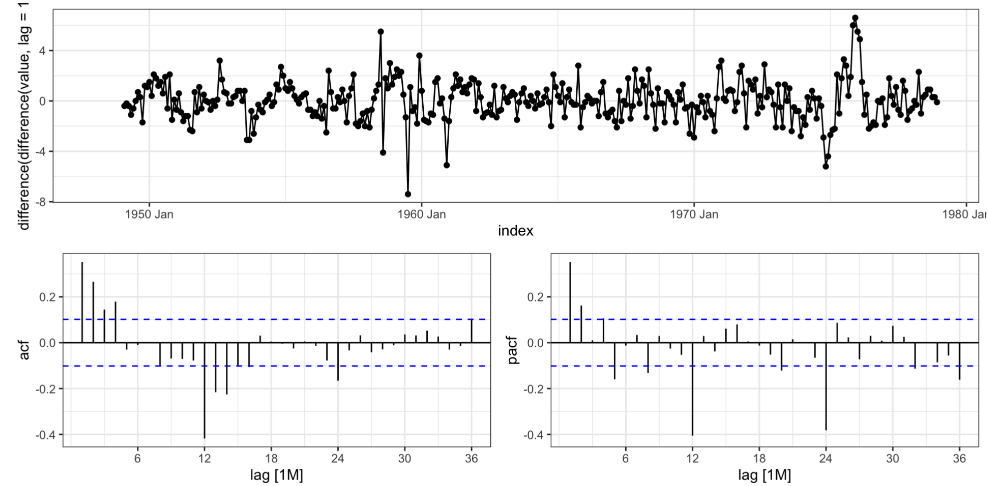


Differencing + Seasonal Differencing

```r
feasts::gg_tsdisplay(
  prodn,
  y = difference(value) |>
    difference(lag=12),
  plot_type = "partial", lag_max=36
)
```

```r
feasts::gg_tsdisplay(
  prodn,
  y = difference(value, lag=12) |>
    difference(),
  plot_type = "partial", lag_max=36
)
```





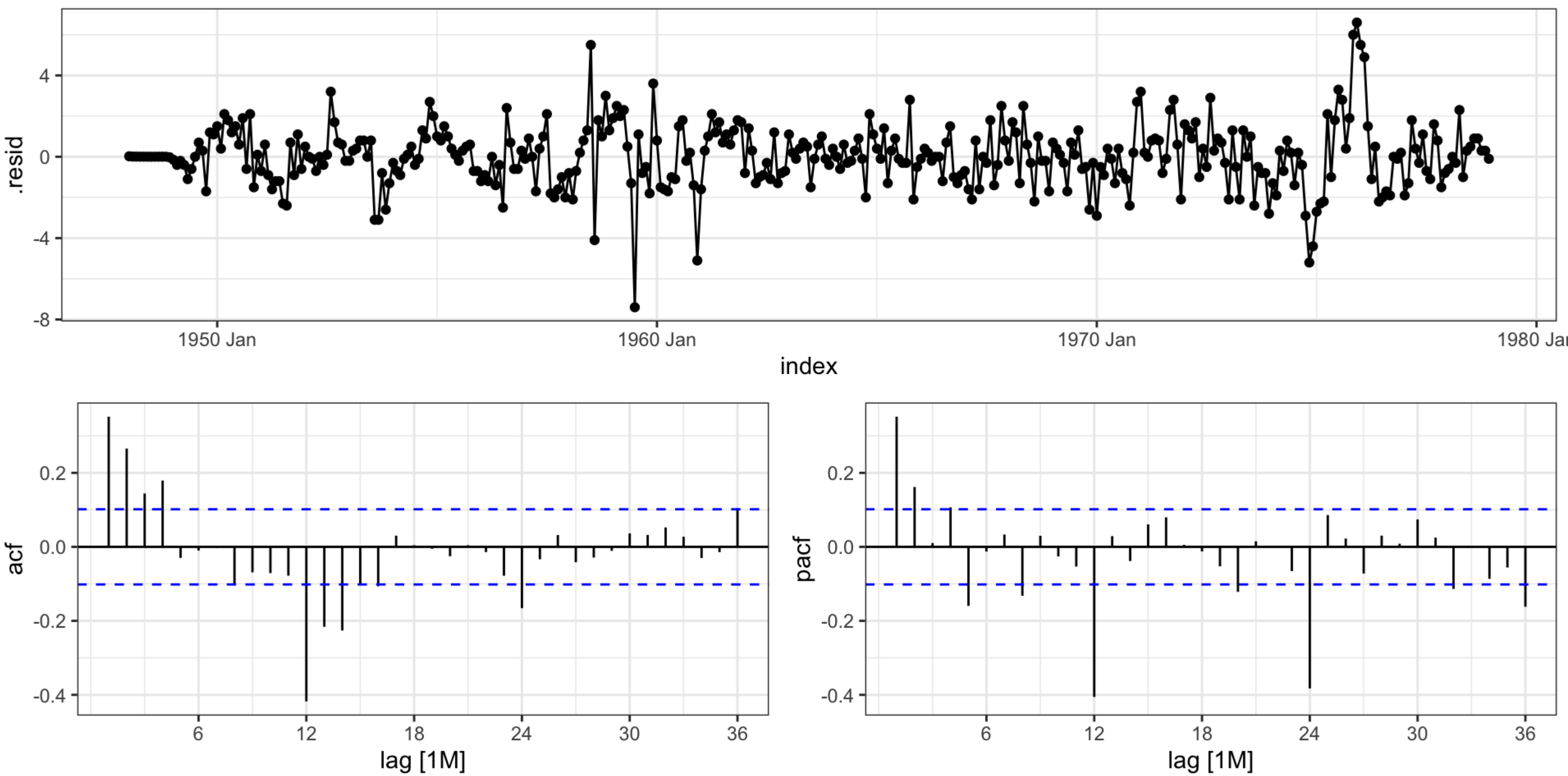
Model 1

```
fr_m = model(
  prodn,
  m1 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,0, period=12)),
)

glance(fr_m)
```

```
# A tibble: 1 × 8
  .model sigma2 log_lik   AIC  AICc   BIC ar_roots   ma_roots
  <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl> <list>     <list>
1 m1       2.52   -675. 1353. 1353. 1356. <cpl [0]> <cpl [0]>
```

Residuals

```
residuals(fr_m) |>
  filter(.model == "m1") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```



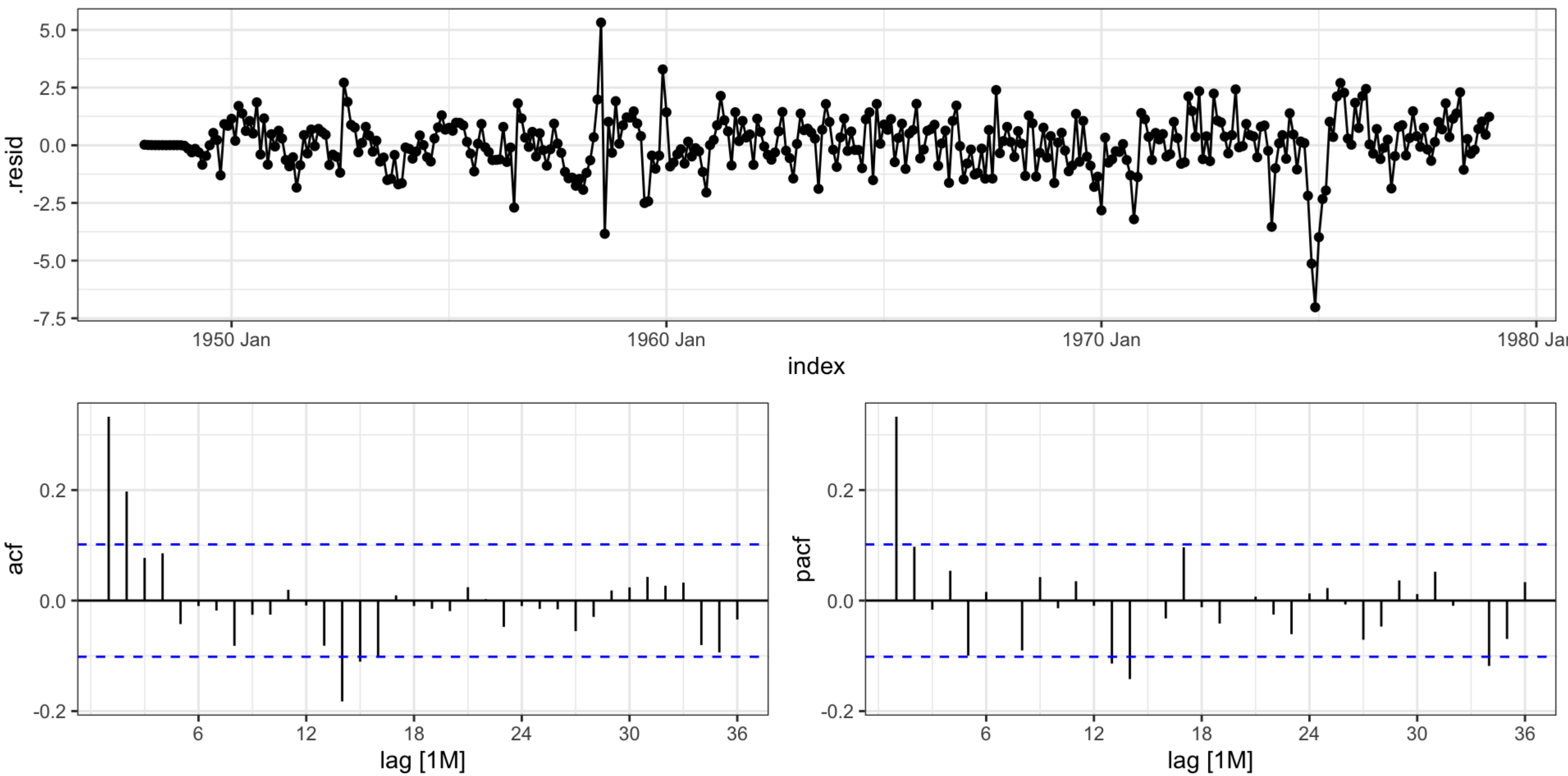
Model 2 - Seasonal MA

```r
fr_m = model(
  prodn,
  m1 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,0, period=12)),
  m2_1 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,1, period=12)),
  m2_2 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,2, period=12)),
  m2_3 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,3, period=12)),
  m2_4 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,4, period=12))
)

glance(fr_m)
```

```
# A tibble: 5 × 8
  .model sigma2 log_lik   AIC  AICc   BIC ar_roots  ma_roots
  <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl> <list>    <list>
1 m1       2.52   -675. 1353. 1353. 1356. <cpl [0]> <cpl [0]>
2 m2_1     1.62   -599. 1203. 1203. 1210. <cpl [0]> <cpl [12]>
3 m2_2     1.62   -599. 1204. 1204. 1216. <cpl [0]> <cpl [24]>
4 m2_3     1.51   -588. 1183. 1183. 1199. <cpl [0]> <cpl [36]>
5 m2_4     1.51   -587. 1185. 1185. 1204. <cpl [0]> <cpl [48]>
```

Residuals - Model 2-3

```r
residuals(fr_m) |>
  filter(.model == "m2_3") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```



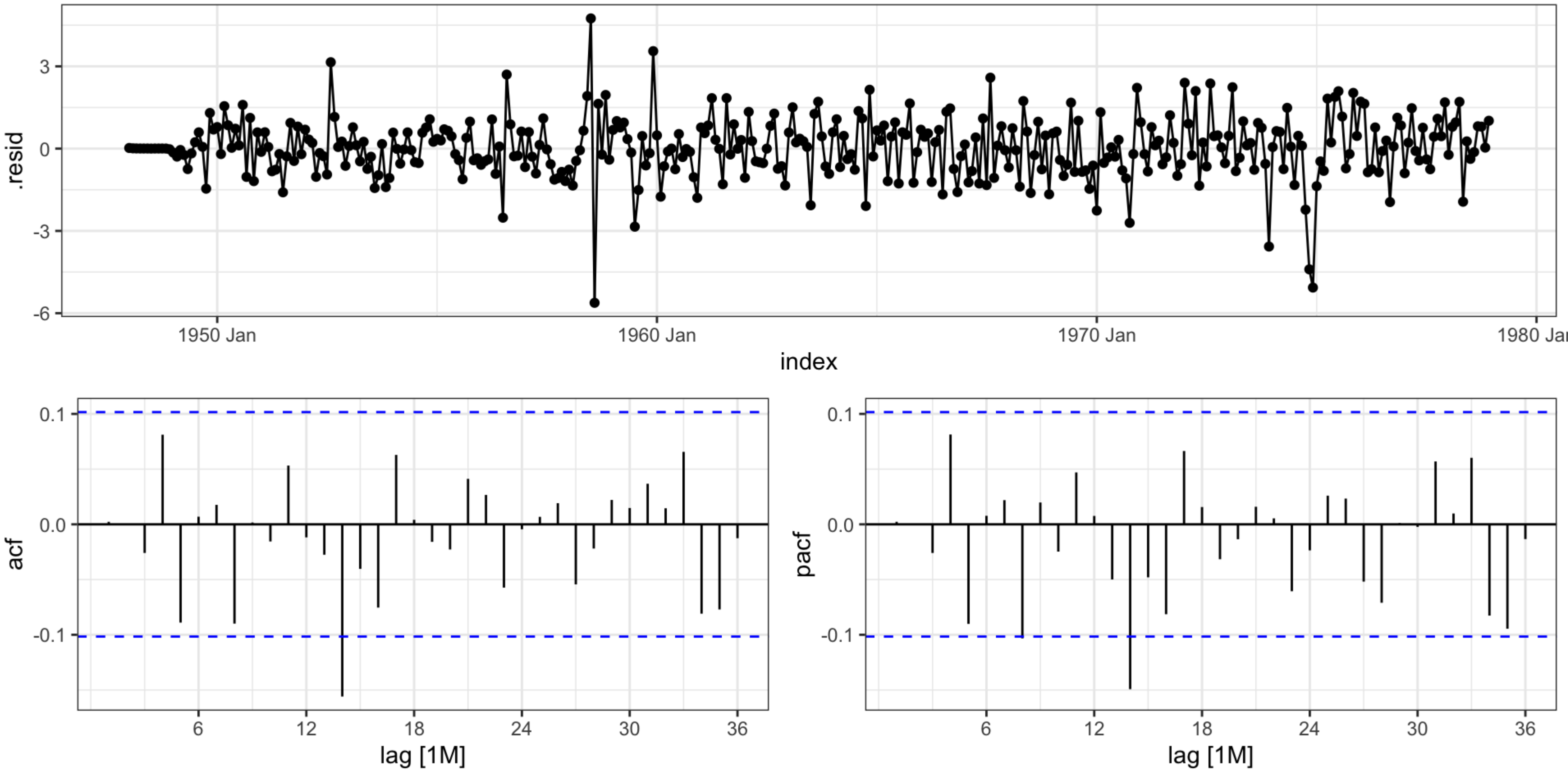
Model 3 - Adding AR

```r
fr_m = model(
  prodn,
  m1 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,0, period=12)),
  m2_3 = ARIMA(value ~ pdq(0,1,0) + PDQ(0,1,3, period=12)),
  m3_1 = ARIMA(value ~ pdq(1,1,0) + PDQ(0,1,3, period=12)),
  m3_2 = ARIMA(value ~ pdq(2,1,0) + PDQ(0,1,3, period=12))
)

glance(fr_m)
```

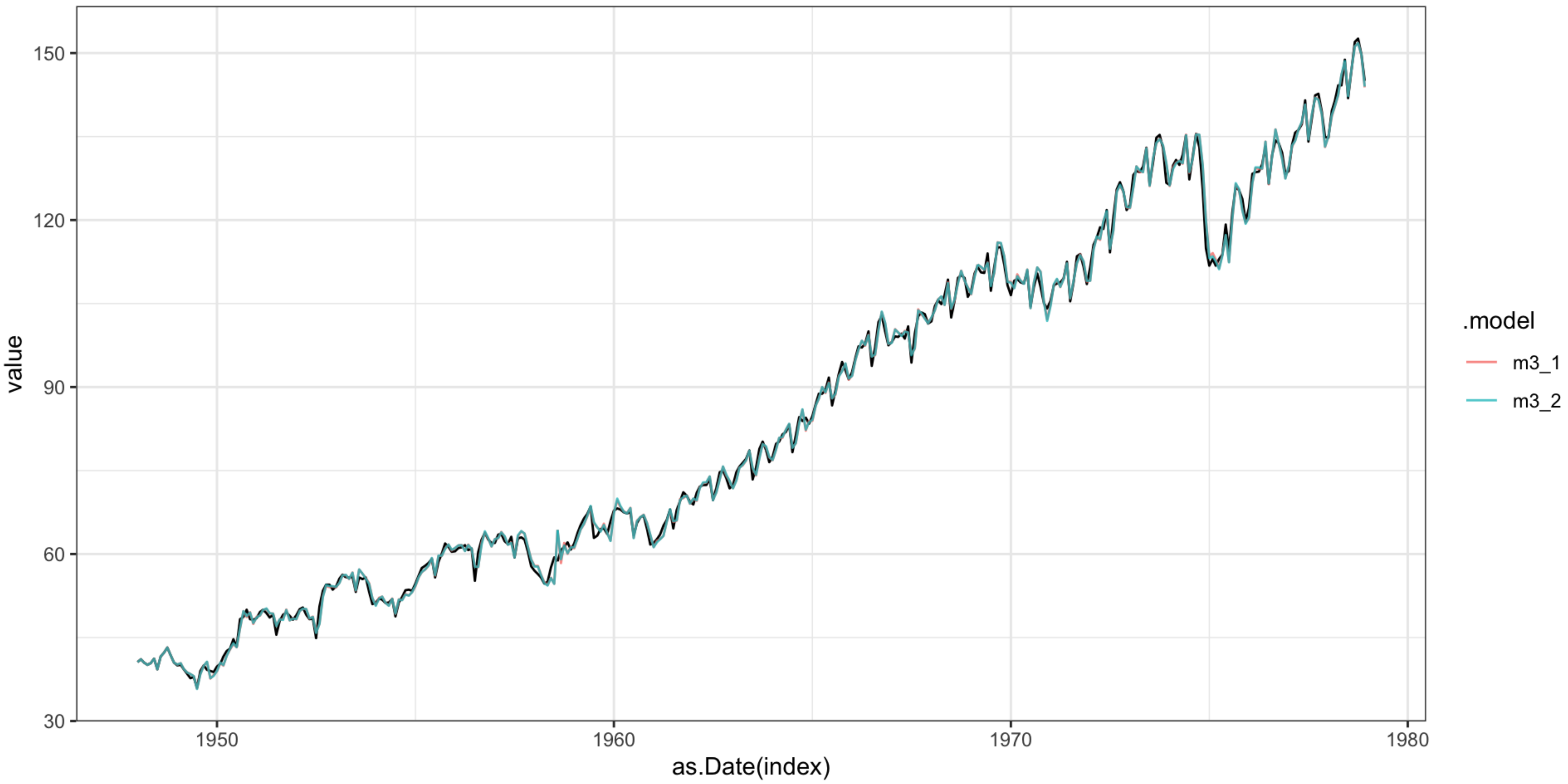
```
# A tibble: 4 × 8
  .model sigma2 log_lik   AIC  AICc   BIC ar_roots  ma_roots
  <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl> <list>    <list>
1 m1       2.52   -675. 1353. 1353. 1356. <cpl [0]> <cpl [0]>
2 m2_3     1.51   -588. 1183. 1183. 1199. <cpl [0]> <cpl [36]>
3 m3_1     1.34   -566. 1142. 1142. 1161. <cpl [1]> <cpl [36]>
4 m3_2     1.33   -564. 1140. 1140. 1163. <cpl [2]> <cpl [36]>
```

Residuals - Model 3-2

```r
residuals(fr_m) |>
  filter(.model == "m3_2") |>
  feasts::gg_tsdisplay(y=.resid, plot_type = "partial", lag_max=36)
```

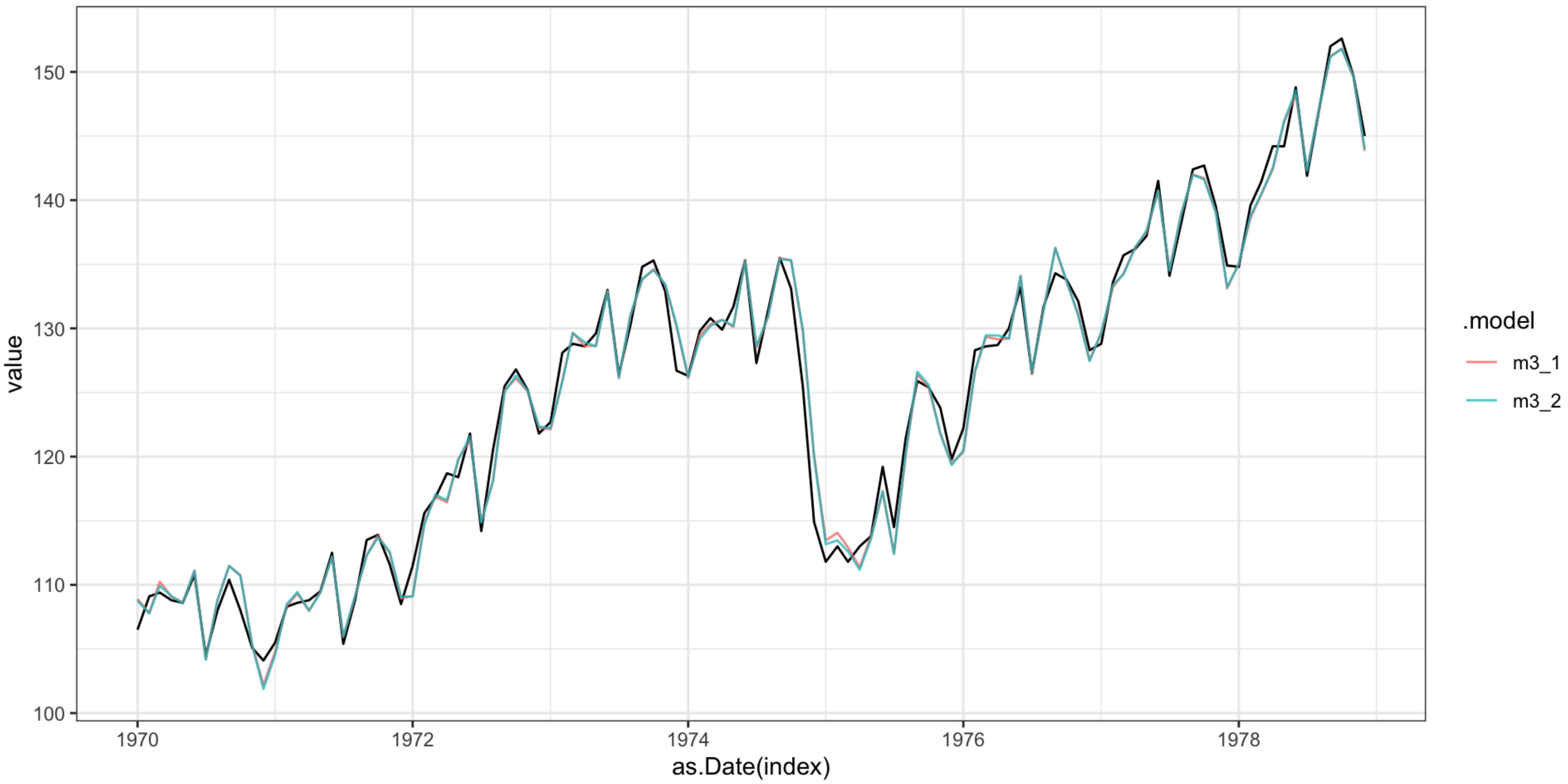


Model Fit



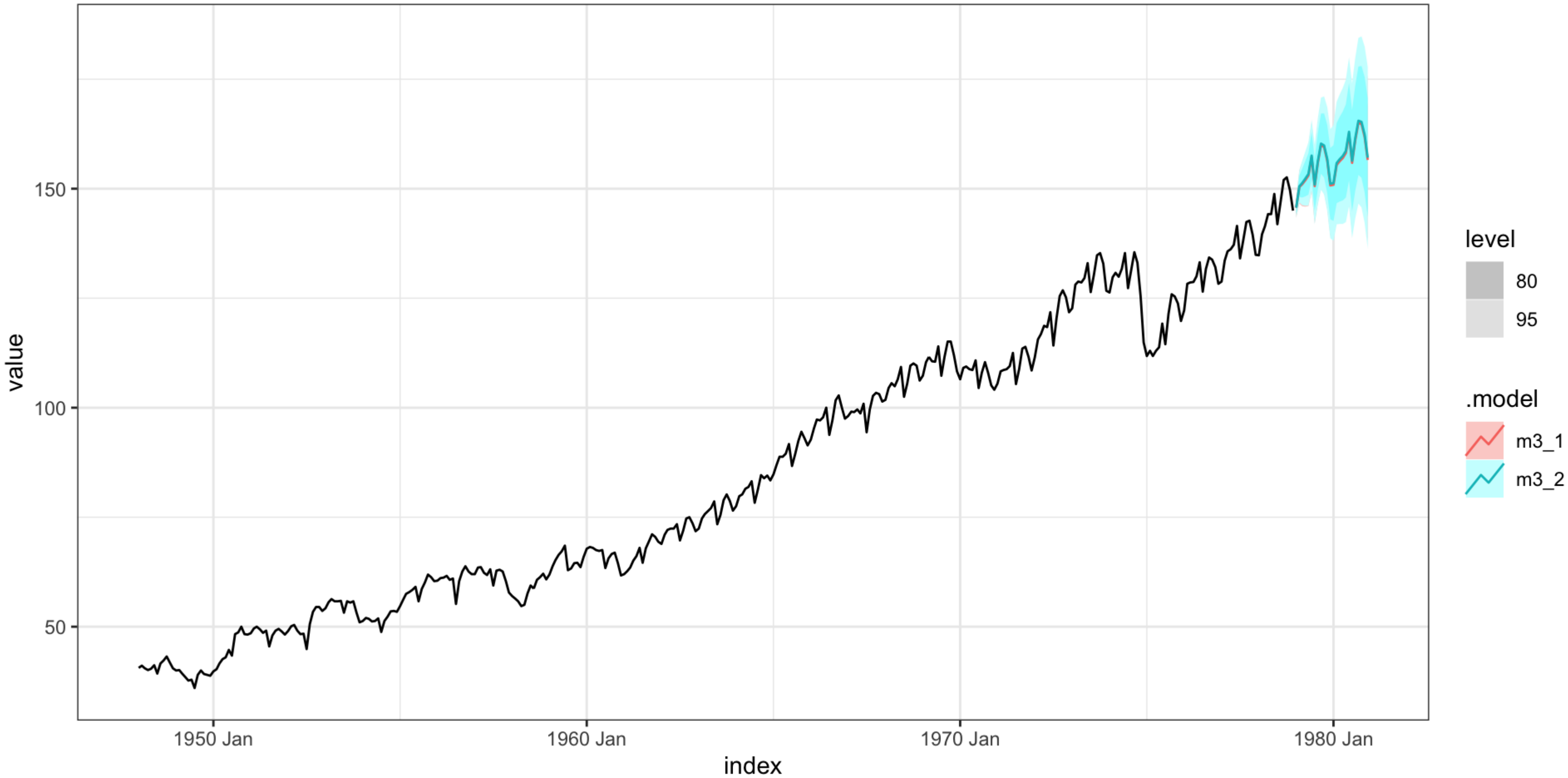
150
120
90
60
30
value
1950
1960
1970
1980
as.Date(index)
.model
m3_1
m3_2

Model Fit (1970-1979)



Model Forecast

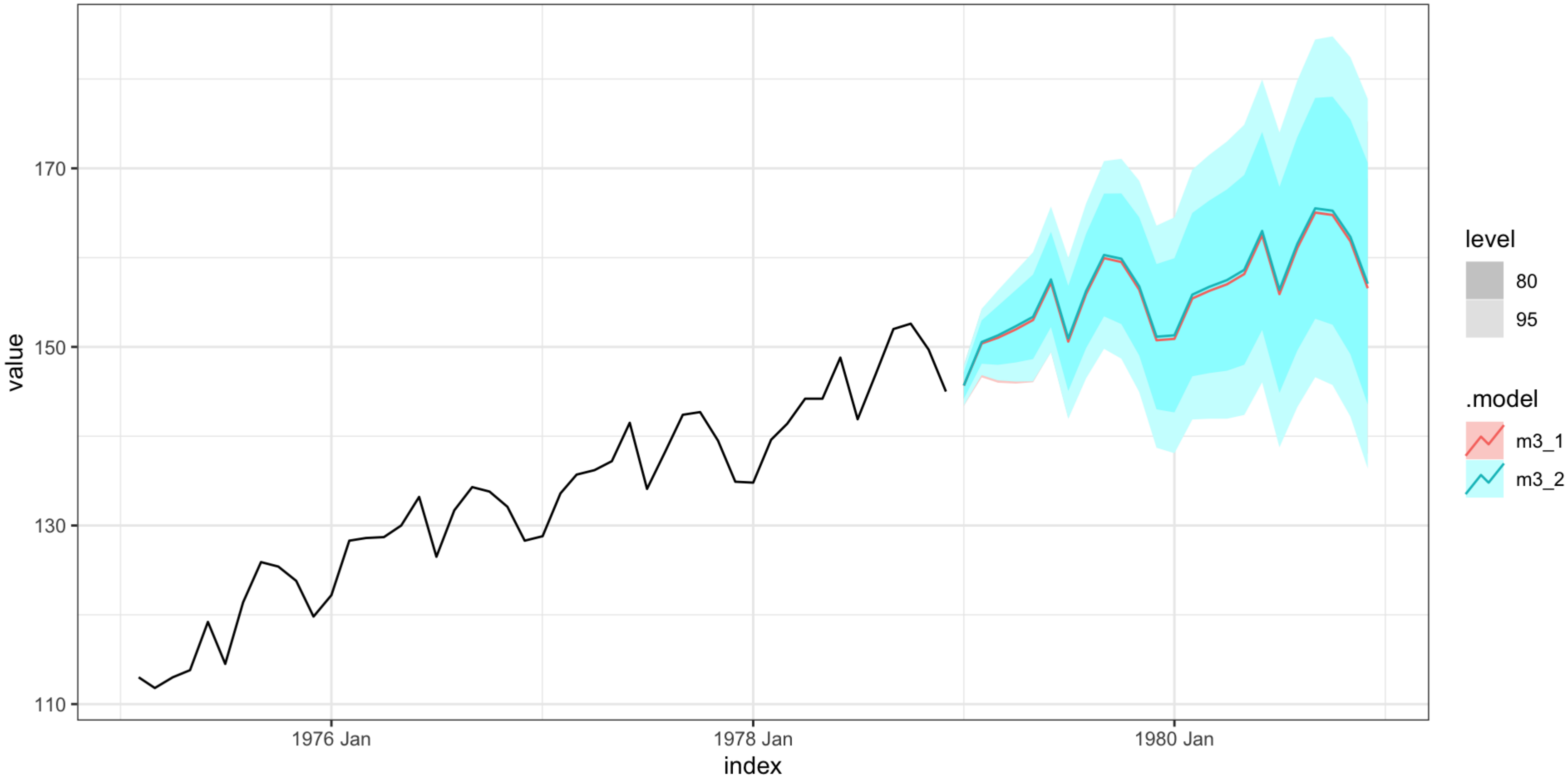
```
fr_m |>
  forecast(h=24) |>
  filter(.model %in% c("m3_1", "m3_2")) |>
  autoplot(
    prodn
  )
```



150
100
50
value
1950 Jan
1960 Jan
1970 Jan
1980 Jan
index
level
80
95
.model
m3_1
m3_2

Model Forecast (1975-1981)

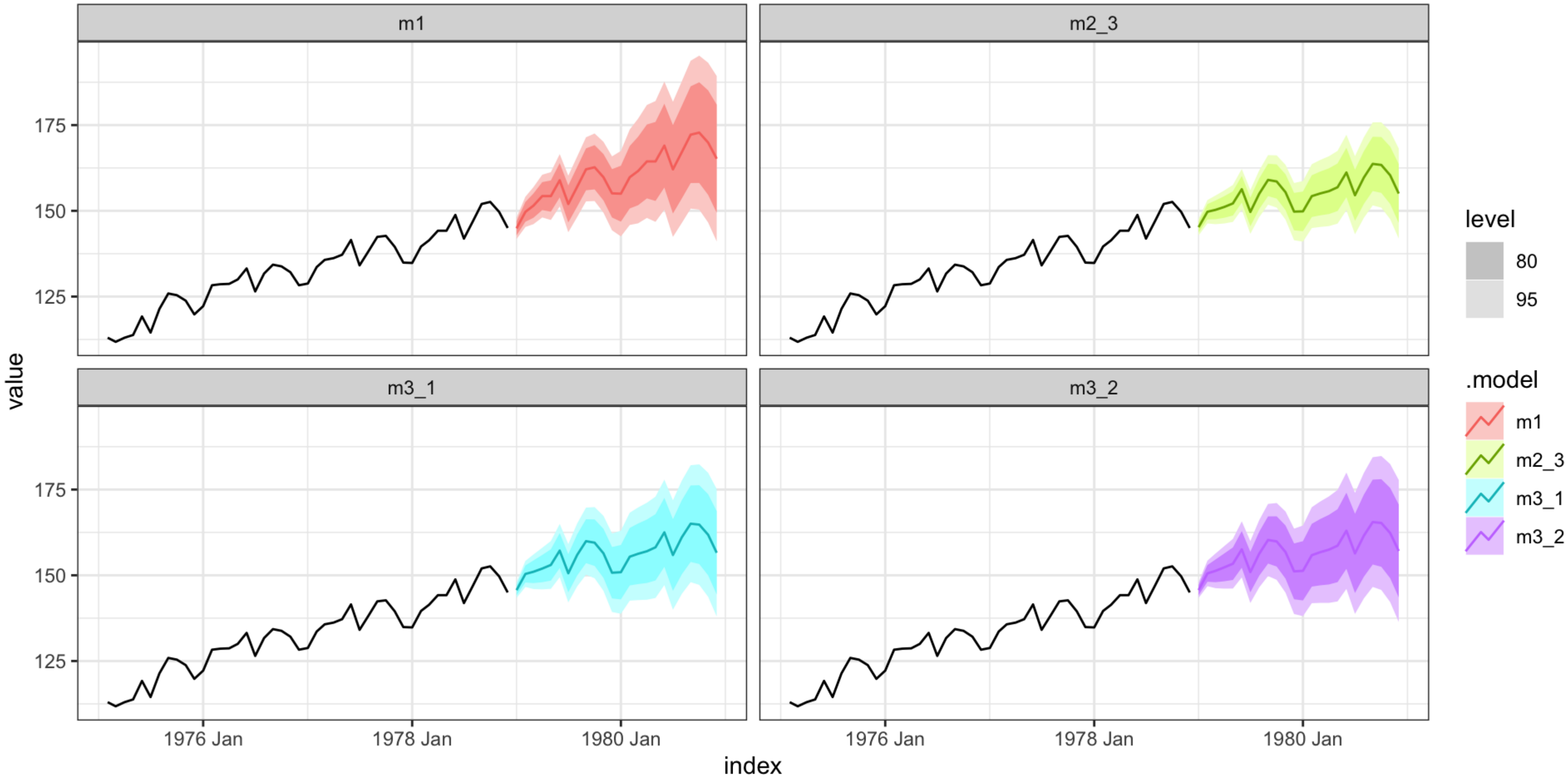
```r
fr_m |>
  forecast(h=24) |>
  filter(.model %in% c("m3_1", "m3_2")) |>
  autoplot(
    prodn |> filter(index > make_yearmonth(1975, 1))
  )
```



170
150
130
110
value
1976 Jan
1978 Jan
1980 Jan
index
level
80
95
.model
m3_1
m3_2

Model Forecast (comparison)

```
fr_m |>
  forecast(h=24) |>
  autoplot(
    prodn |> filter(index > make_yearmonth(1975, 1))
  ) +
    facet_wrap(~.model)
```



m1
m2_3
m3_1
m3_2
value
175
150
125
1976 Jan
1978 Jan
1980 Jan
index
level
80
95
.model
m1
m2_3
m3_1
m3_2

Auto ARIMA - Model Fit

```r
model(prodn, ARIMA(value)) |>
  report()
```

```
Series: value
Model: ARIMA(2,0,1)(0,1,1)[12] w/ drift

Coefficients:
         ar1      ar2      ma1     sma1  constant
      1.6696  -0.6944  -0.4166  -0.6684    0.0857
s.e.  0.0999   0.0973   0.1286   0.0352    0.0125

sigma^2 estimated as 1.398:  log likelihood=-573.59
AIC=1159.18   AICc=1159.42   BIC=1182.5
```

Exercise - Cortecosteroid Drug Sales

Monthly cortecosteroid drug sales in Australia from 1992 to 2008.

```r
data(h02, package="fpp")
h02 = as_tsibble(h02)

feasts::gg_tsdisplay(h02, y=value, plot_type = "partial", lag_max=36)
```