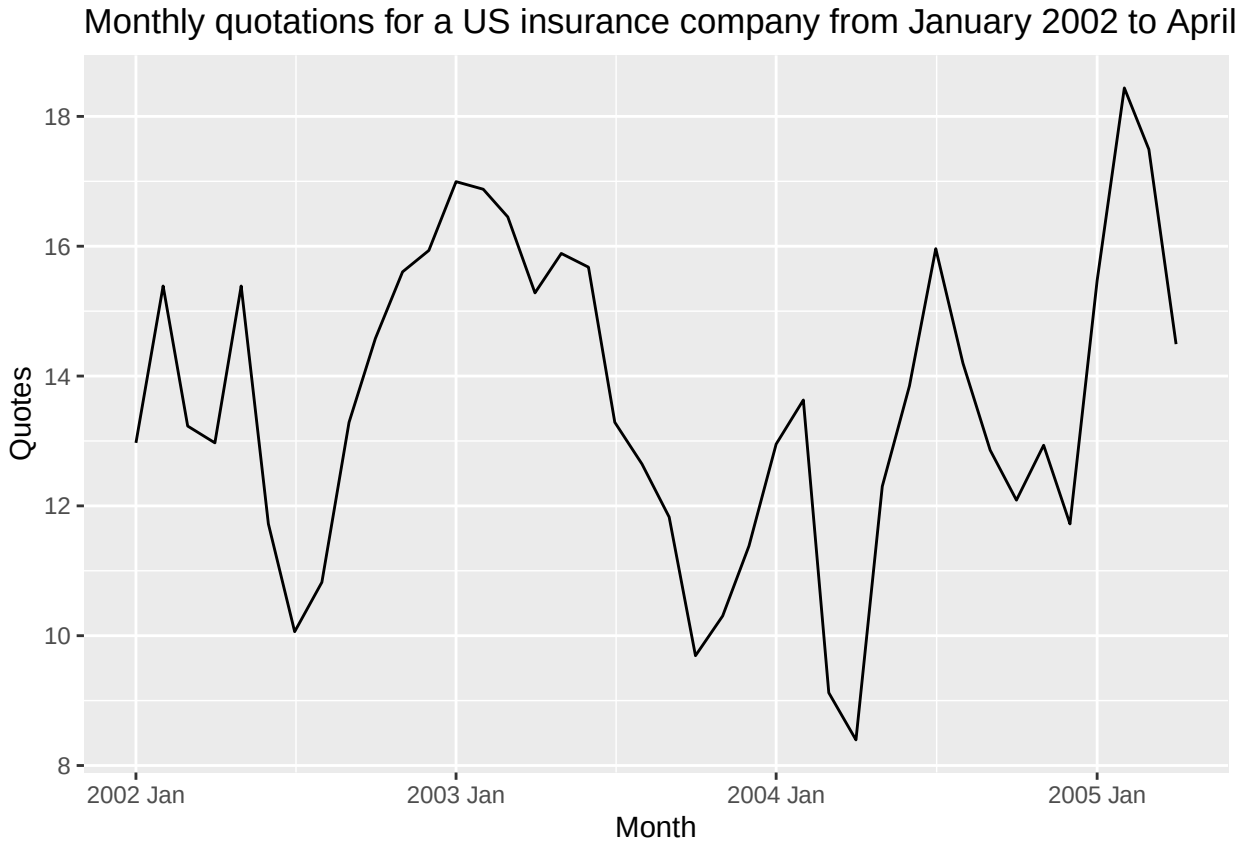
Time Series Project

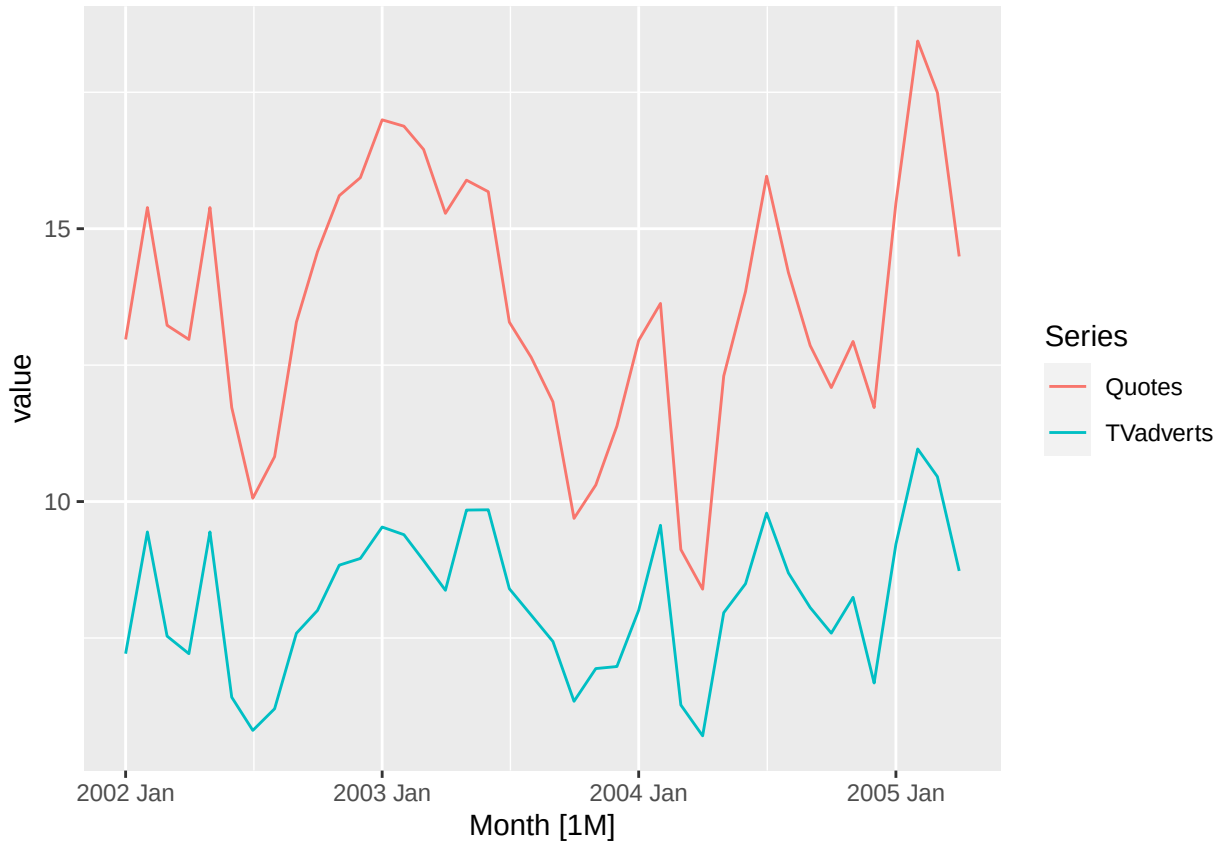
2023-11-17

```
library(fpp3)
library(prophet)
library(fable.prophet)
```

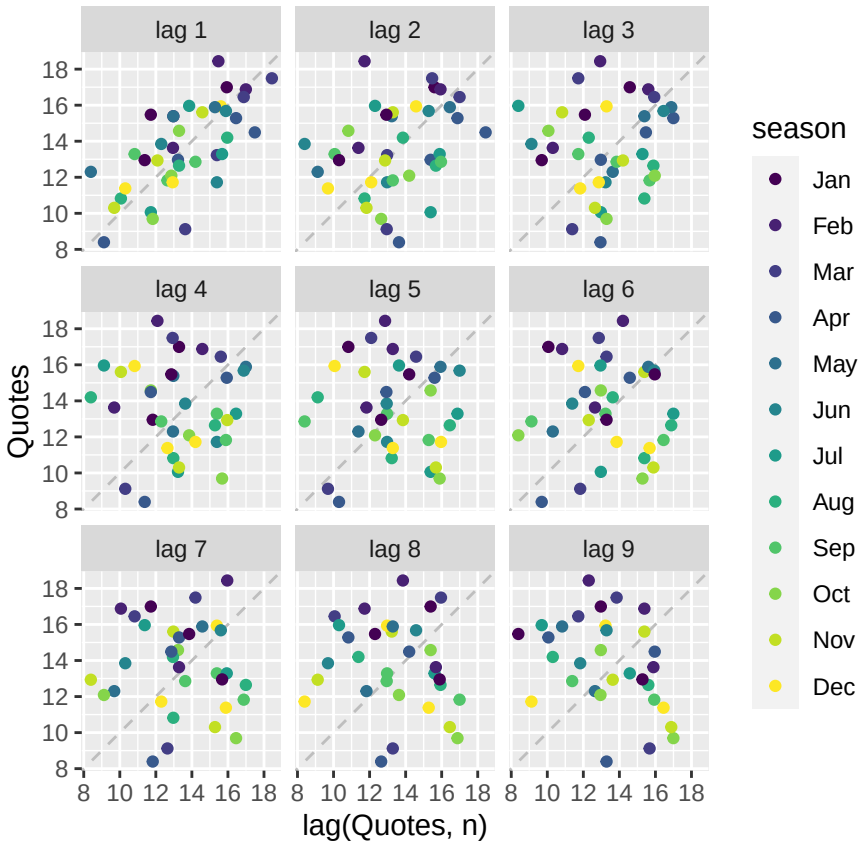
```
insurance |> autoplot(Quotes)+
  labs(y = "Quotes", x = "Month",
       title = "Monthly quotations for a US insurance company from January 2002 to April 2005")
```



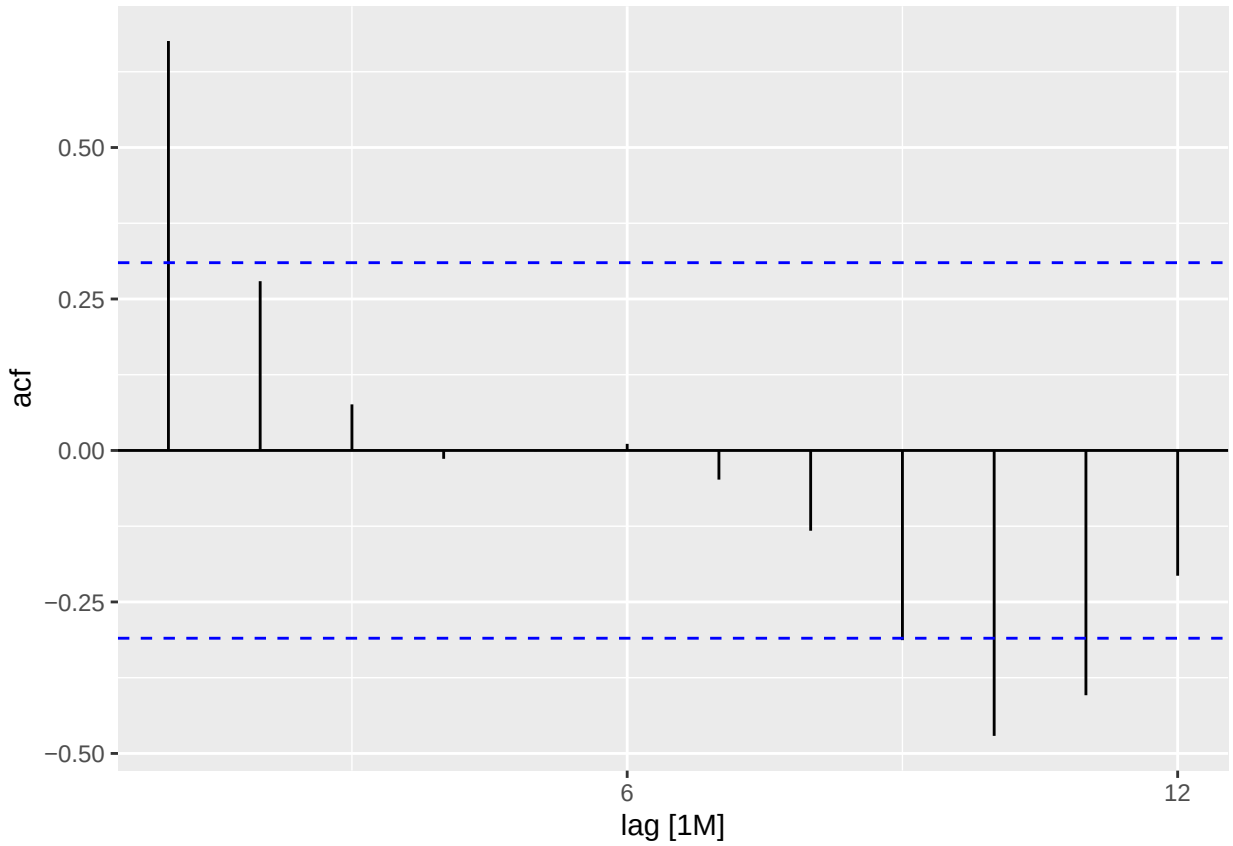

```
insurance |>
  pivot_longer(c(Quotes, TVadverts), names_to="Series") |>
  autoplot(value)
```



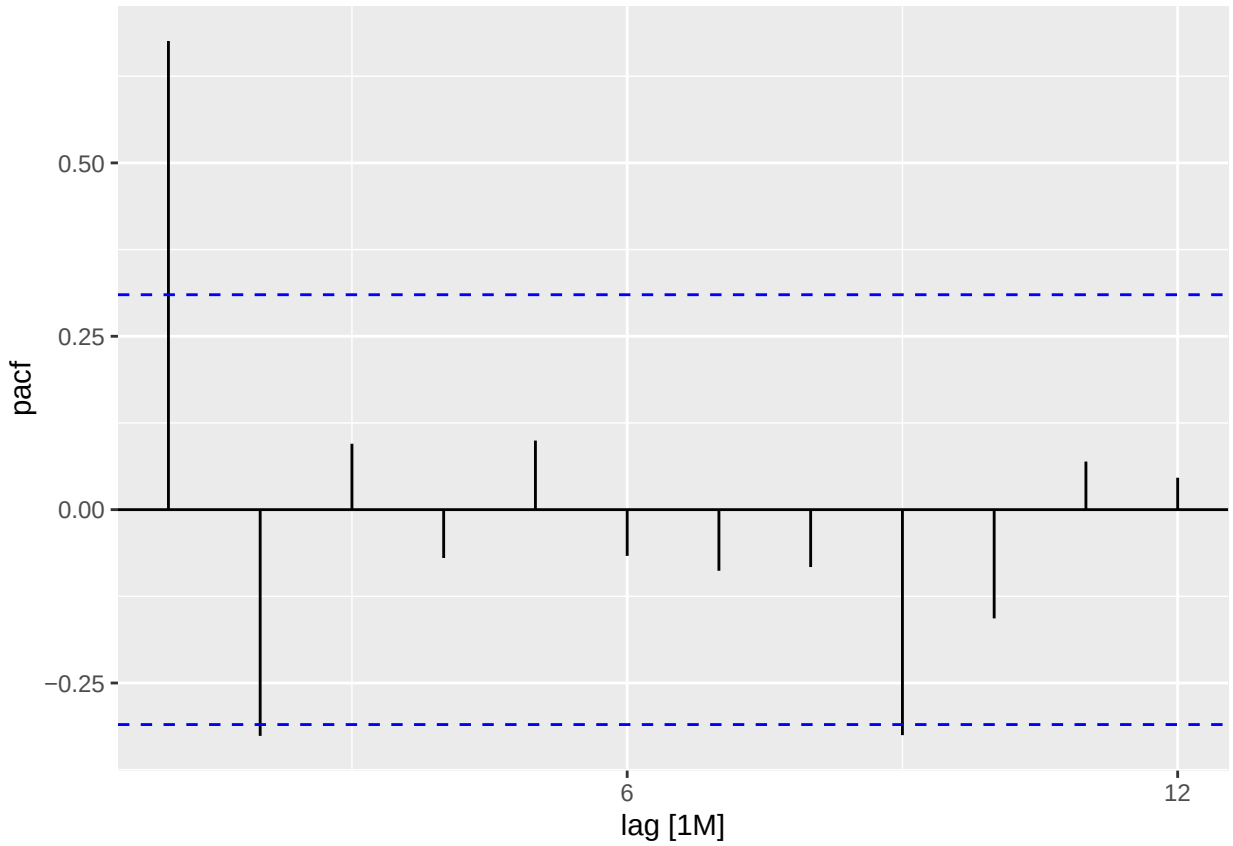

```
insurance |> gg_lag(Quotes, geom='point')
```




```
insurance |> ACF(Quotes, lag_max = 12) |> autoplot()
```

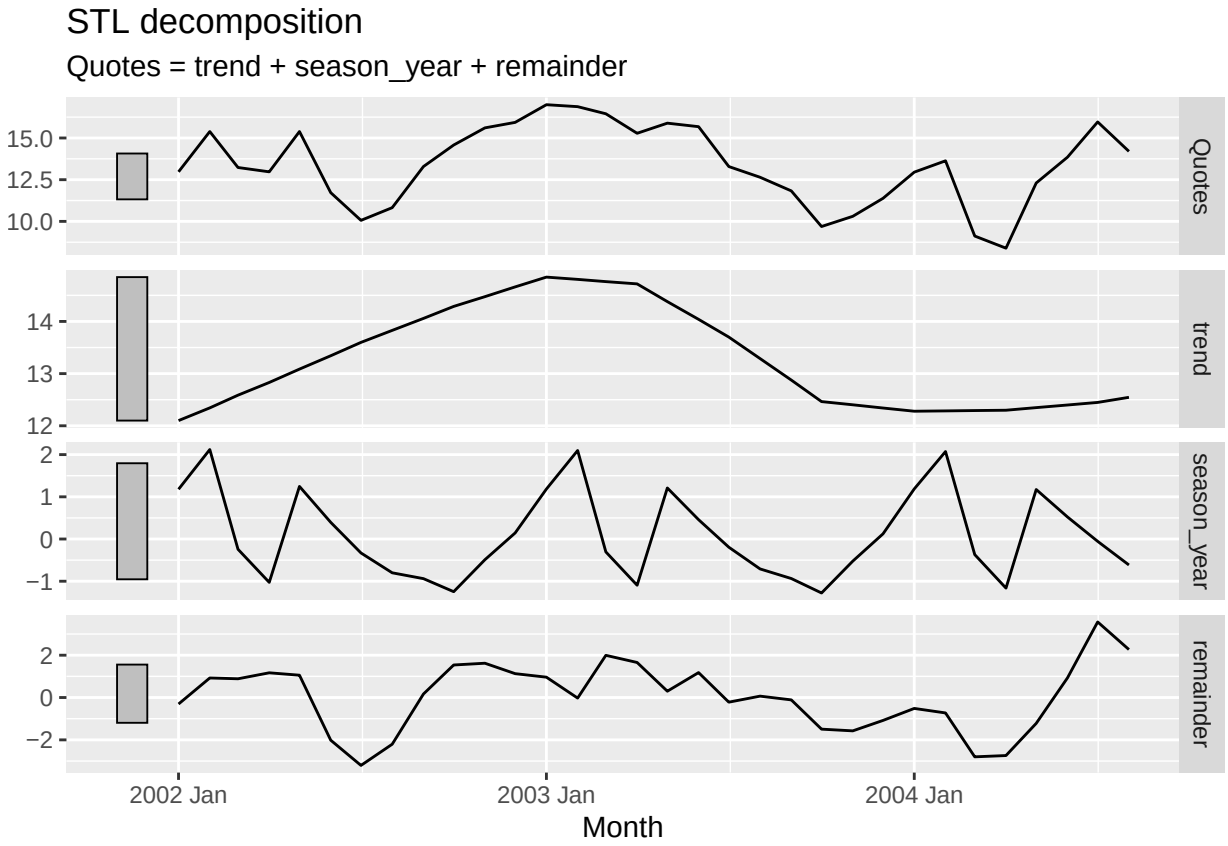



```
insurance |> PACF(Quotes, lag_max = 12) |> autoplot()
```




```
Train <- insurance[0:32,]
Test <- insurance[33:40,]
```

```
ins_stl <- Train |>
  model(stl = STL(Quotes))

ins_stl |>
  components() |>
  autoplot()
```



Fitting Models

```
fit1 <-  Train |>  model(
        lm     =  TSLM(Quotes),
        lm2    =  TSLM(Quotes ~ TVadverts),
        lm3    =  TSLM(Quotes ~ lag(TVadverts)),
        lm4    =  TSLM(Quotes ~ TVadverts + lag(TVadverts)),
        ARIMA  =  ARIMA(Quotes),
        ARIMA2 =  ARIMA(Quotes ~ TVadverts),
        ARIMA3 =  ARIMA(Quotes ~ lag(TVadverts)),
        ARIMA4 =  ARIMA(Quotes ~ TVadverts + lag(TVadverts)),
        Mean   =  MEAN(Quotes),
        SNaive =  SNAIVE(Quotes),
        Naive  =  NAIVE(Quotes),
        Drift  =  RW(Quotes ~ drift()),
        ETS    =  ETS(Quotes),
        NNET   =  NNETAR(Quotes),
        Prophet = prophet(Quotes)
)
```

```
## n.changepoints greater than number of observations. Using 24
```

```
## code for methods in class "Rcpp_model_base" was not checked for suspicious field assignments (recomme
## code for methods in class "Rcpp_model_base" was not checked for suspicious field assignments (recomme
```

```
## code for methods in class "Rcpp_stan_fit" was not checked for suspicious field assignments (recommend
## code for methods in class "Rcpp_stan_fit" was not checked for suspicious field assignments (recommend
```

```r
acc1 <- accuracy(fit1) |> arrange(RMSE)
acc1
```

```
## # A tibble: 15 x 10
##    .model .type        ME   RMSE    MAE      MPE   MAPE    MASE   RMSSE     ACF1
##    <chr>  <chr>     <dbl>  <dbl>  <dbl>    <dbl>  <dbl>   <dbl>   <dbl>    <dbl>
##  1 NNET   Trai~ -6.15e- 5 0.0261 0.0165 -3.38e-3  0.119 0.00487 0.00683 -0.555
##  2 ARIMA4 Trai~ -2.04e- 2 0.510  0.406  -1.73e-2  3.17  0.120   0.133    0.00799
##  3 ARIMA2 Trai~ -1.75e- 2 0.536  0.416  -1.38e-1  3.23  0.122   0.140    0.0227
##  4 Proph~ Trai~ -7.94e- 5 0.723  0.565  -3.83e-1  4.24  0.166   0.189    0.319
##  5 lm4    Trai~ -1.11e-16 0.926  0.804  -5.51e-1  6.24  0.237   0.242    0.744
##  6 lm2    Trai~  0        0.948  0.842  -5.79e-1  6.50  0.248   0.248    0.653
##  7 ARIMA3 Trai~ -1.35e- 2 1.48   1.18   -1.69e+0  9.42  0.348   0.388    0.0336
##  8 ARIMA  Trai~  2.25e- 2 1.56   1.21   -1.40e+0  9.56  0.358   0.409   -0.0168
##  9 ETS    Trai~  3.83e- 2 1.83   1.45   -8.65e-1 11.7   0.428   0.478    0.0618
## 10 Drift  Trai~  7.02e-16 1.86   1.50   -1.20e+0 12.1   0.442   0.486    0.0626
## 11 Naive  Trai~  3.96e- 2 1.86   1.50   -8.92e-1 12.1   0.442   0.486    0.0626
## 12 lm3    Trai~ -2.65e-16 2.09   1.75   -2.82e+0 14.2   0.517   0.546    0.395
## 13 lm     Trai~  0        2.33   1.92   -3.46e+0 15.6   0.566   0.609    0.689
## 14 Mean   Trai~  5.90e-16 2.33   1.92   -3.46e+0 15.6   0.566   0.609    0.689
## 15 SNaive Trai~ -9.17e- 1 3.82   3.40   -1.26e+1 28.8   1       1        0.755
```

```r
print(c("The best mdoel on Train is: ", acc1$.model[1]))
```

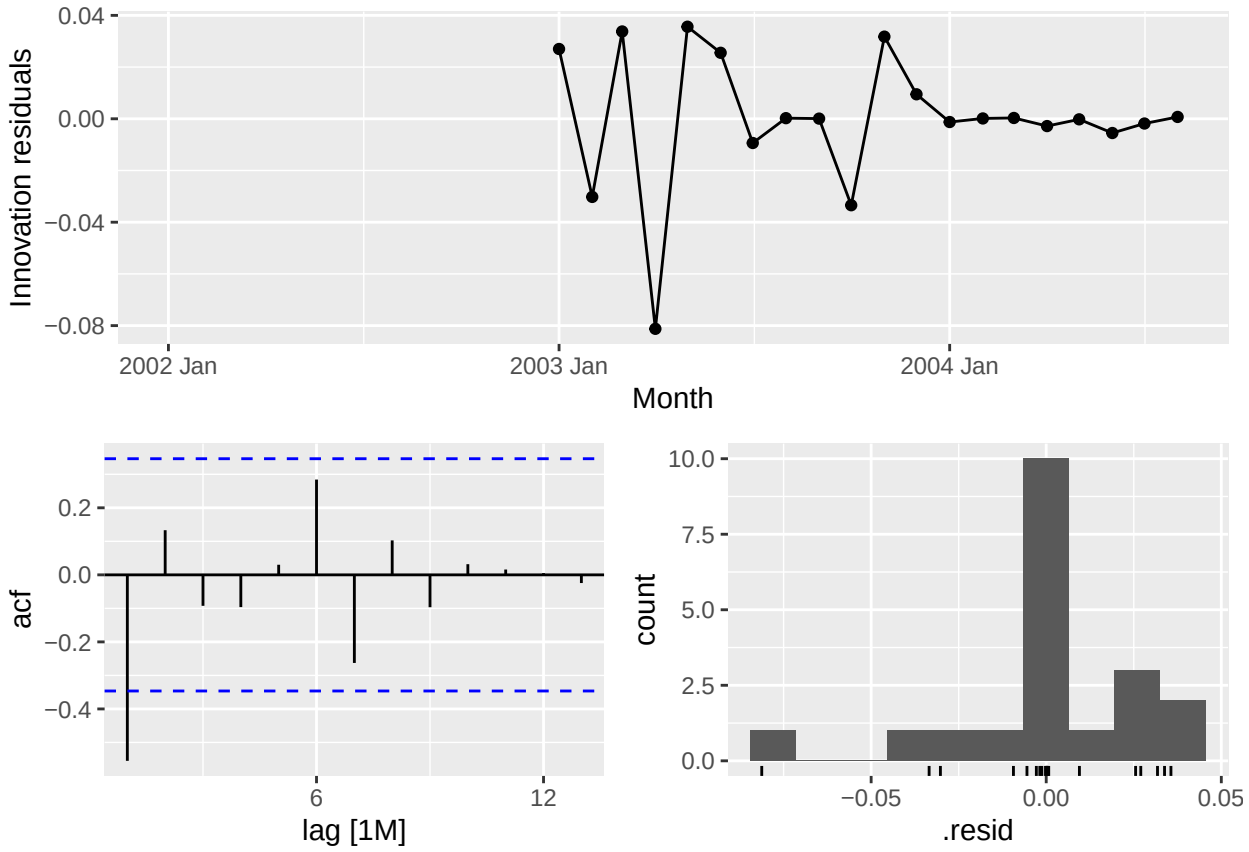
```
## [1] "The best mdoel on Train is: " "NNET"
```

```r
fit1[,acc1$.model[1]] |>  gg_tsresiduals()
```

```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
```

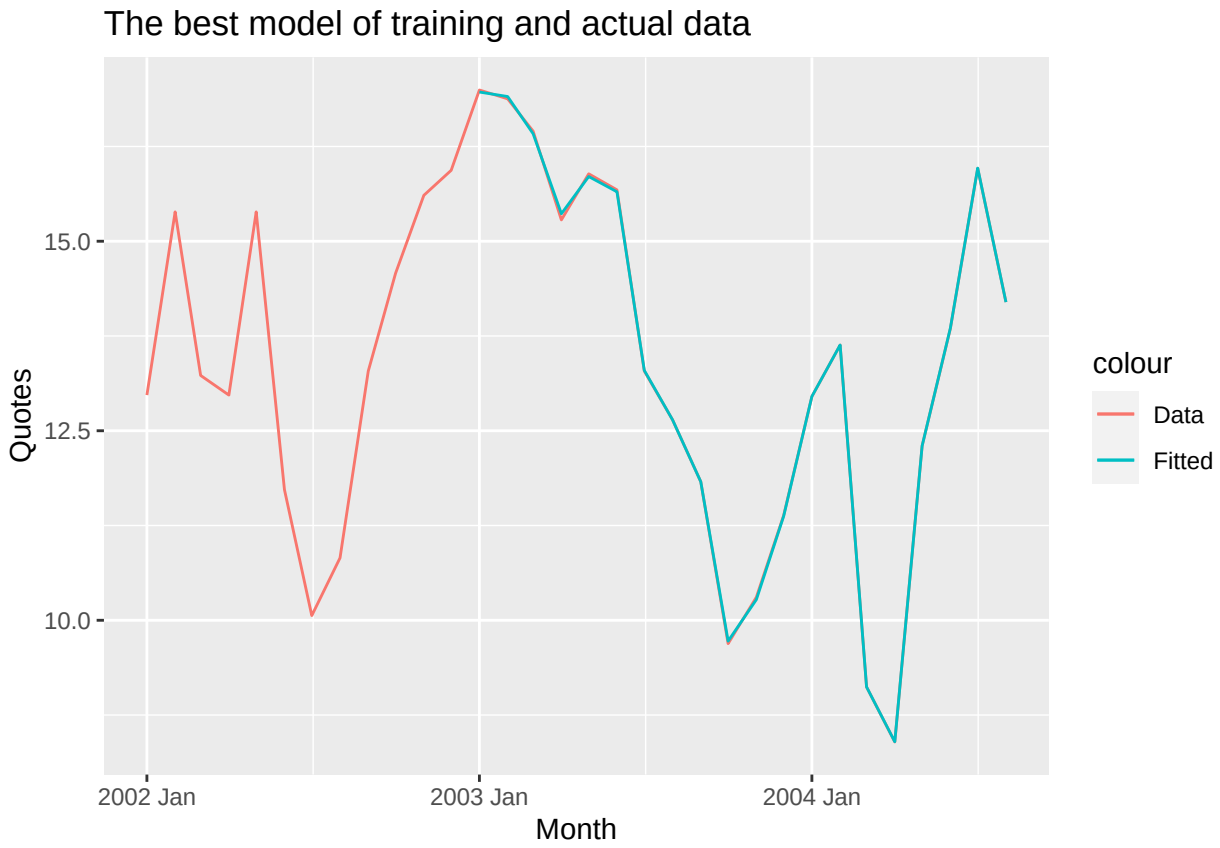
```
## Warning: Removed 12 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 12 rows containing non-finite values (`stat_bin()`).
```




```
# augment(fit1) |> filter(.model == acc1$.model[1]) |>
#   features(.resid, ljung_box, lag=12, dof=0)
#
# augment(fit1) |> filter(.model == acc1$.model[1]) |>
#   features(.innov, box_pierce, lag = 12,  dof = 0)


augment(fit1) |> filter(.model == acc1$.model[1]) |>
  ggplot(aes(x = Month)) +
  geom_line(aes(y = Quotes , colour = "Data")) +
  geom_line(aes(y = .fitted, colour = "Fitted"))+
  ggtitle("The best model of training and actual data")
```

```
## Warning: Removed 12 rows containing missing values (`geom_line()`).
```




```
fc1 <- fit1 |> forecast(new_data = Test)


accT1 <- accuracy(fc1, Test) |> arrange(RMSE)
accT1
```

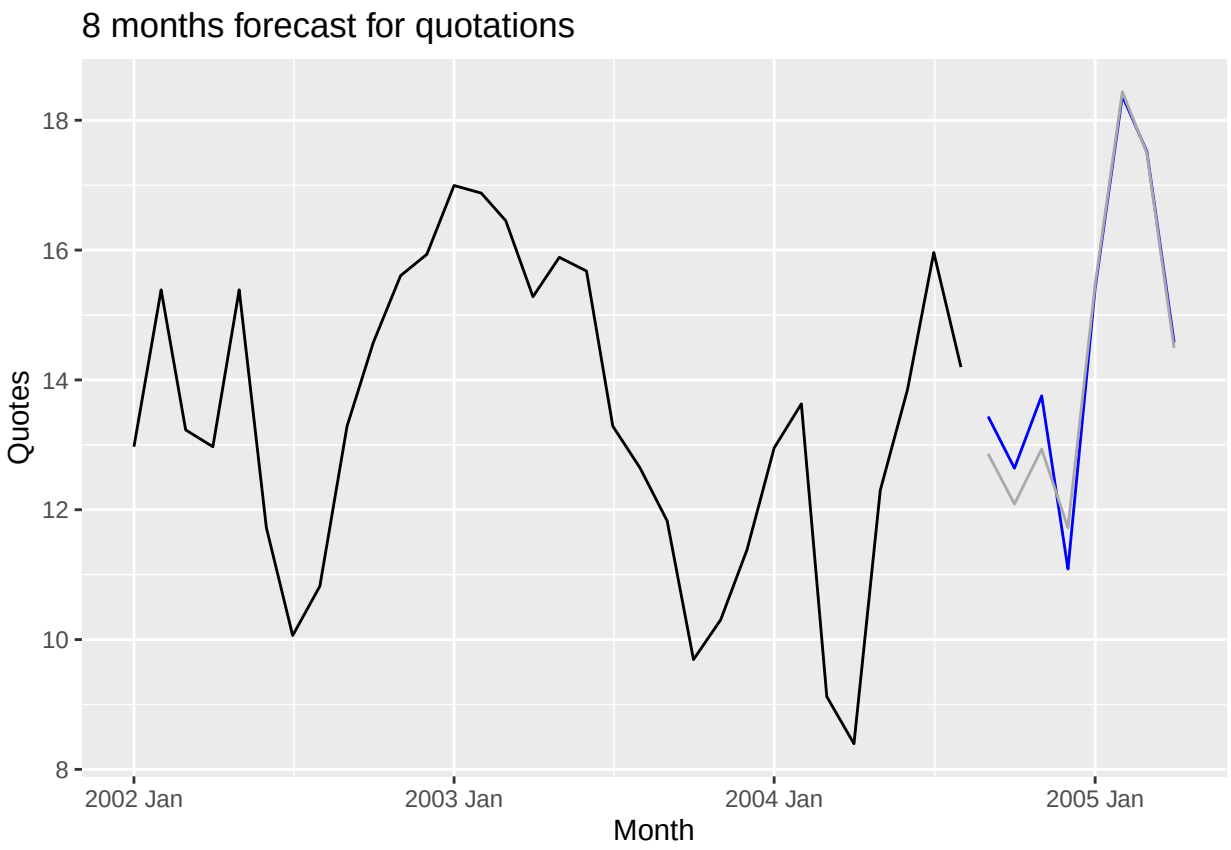
```
## # A tibble: 15 x 10
##    .model  .type      ME   RMSE   MAE     MPE  MAPE  MASE RMSSE   ACF1
##    <chr>   <chr>   <dbl>  <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl>  <dbl>
##  1 lm2     Test  -0.163   0.465 0.352  -1.25   2.78   NaN   NaN  0.104
##  2 lm4     Test  -0.153   0.475 0.435  -1.10   3.23   NaN   NaN  0.240
##  3 ARIMA2  Test   0.451   0.935 0.769   2.40   4.93   NaN   NaN  0.693
##  4 ARIMA4  Test   0.650   0.943 0.713   4.00   4.51   NaN   NaN  0.657
##  5 lm3     Test   0.609   2.24  1.87    2.19  12.4    NaN   NaN  0.213
##  6 Drift   Test   0.0617  2.29  1.96   -2.02  13.5    NaN   NaN  0.551
##  7 ETS     Test   0.240   2.36  2.04   -0.831 13.9    NaN   NaN  0.568
##  8 Naive   Test   0.240   2.36  2.04   -0.829 13.9    NaN   NaN  0.568
##  9 Mean    Test   1.04    2.57  2.04    4.86  13.1    NaN   NaN  0.568
## 10 lm      Test   1.04    2.57  2.04    4.86  13.1    NaN   NaN  0.568
## 11 ARIMA   Test   1.28    2.59  1.97    6.66  12.5    NaN   NaN  0.542
## 12 ARIMA3  Test   1.45    3.01  2.27    7.56  14.3    NaN   NaN  0.624
## 13 NNET    Test  -1.66    4.06  3.45  -15.4   25.5    NaN   NaN  0.480
## 14 SNaive  Test   3.52    4.34  3.52   22.9   22.9    NaN   NaN  0.533
## 15 Prophet Test   4.33   10.5   8.64   31.9   65.6    NaN   NaN -0.210
```

```
print(c("The best mdoel on Test is: ", accT1$.model[1]))
```

```
## [1] "The best mdoel on Test is: " "lm2"
```

```
fc1 |> filter(.model == accT1$.model[1]) |>
  autoplot(Train, level = NULL) +
  labs(y = "Quotes", x = "Month",
  title = "8 months forecast for quotations") +
    autolayer(Test, colour = "darkgray")
```

```
## Plot variable not specified, automatically selected `.vars = Quotes`
```



Fitting Combination Models

```
fit2 <- fit1 |>  mutate(
    Combination1 = (lm2 + lm4)/2,
    Combination2 = (lm2 + ARIMA2)/2,
    Combination3 = (lm2 + ARIMA4)/2,
    Combination4 = (lm4 + ARIMA2)/2,
    Combination5 = (lm4 + ARIMA4)/2,
    Combination6 = (ARIMA2 + ARIMA4)/2,
    Combination7 = (NNET + ARIMA4)/2,
    Combination8 = (NNET + ARIMA2)/2,
    Combination8 = (NNET + Prophet)/2,
    Combination8 = (ARIMA4 + ARIMA2)/2,
    Combination8 = (ARIMA4 + Prophet)/2,
    Combination8 = (ARIMA2 + Prophet)/2,
  )

acc2 <- accuracy(fit2) |> arrange(RMSE)
```

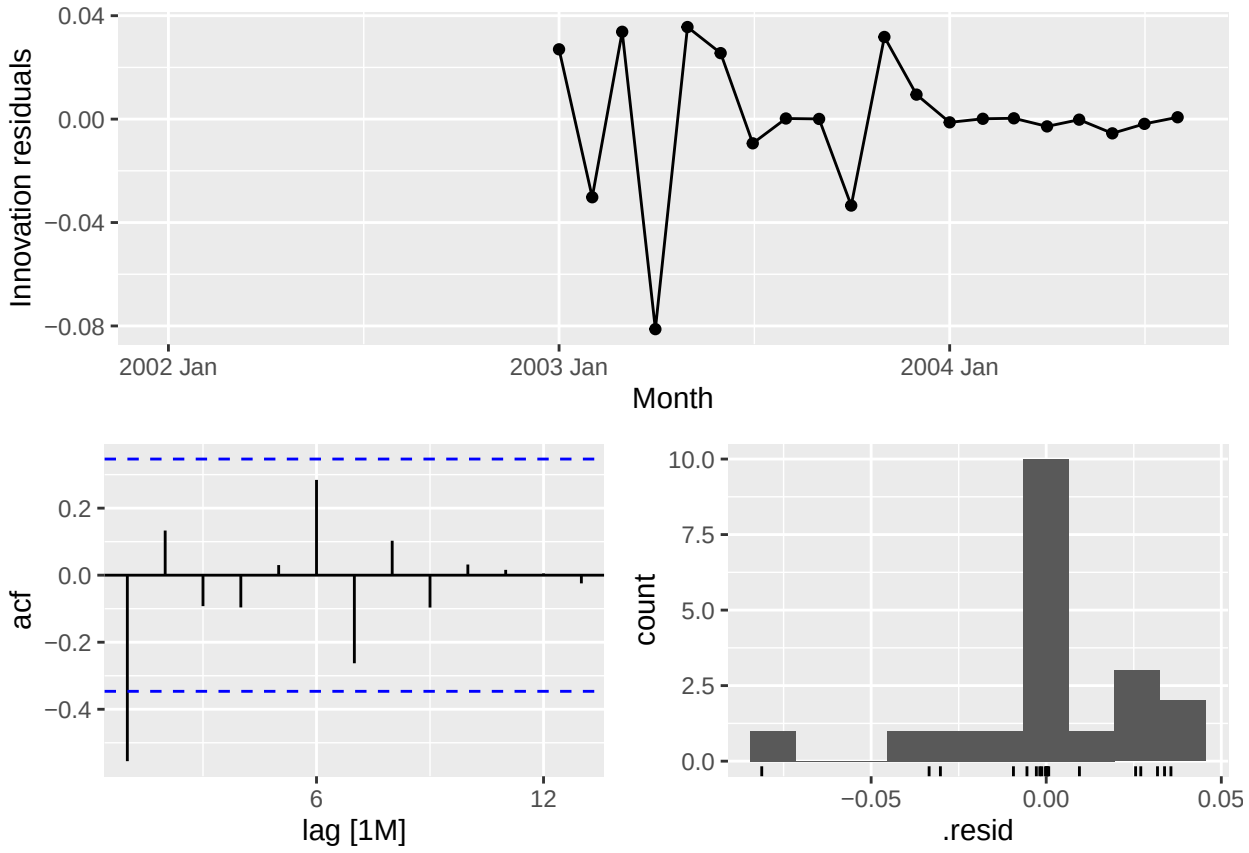
```
acc2
```

```
## # A tibble: 23 x 10
##    .model   .type       ME   RMSE    MAE      MPE  MAPE    MASE   RMSSE      ACF1
##    <chr>    <chr>    <dbl>  <dbl>  <dbl>    <dbl> <dbl>   <dbl>   <dbl>     <dbl>
##  1 NNET     Trai~ -6.15e-5 0.0261 0.0165 -0.00338 0.119 0.00487 0.00683 -0.555
##  2 Combina~ Trai~ -3.34e-2 0.289  0.234  -0.155   1.82  0.0689  0.0756   0.0559
##  3 Combina~ Trai~ -8.78e-3 0.481  0.385  -0.261   2.95  0.113   0.126    0.156
##  4 ARIMA4   Trai~ -2.04e-2 0.510  0.406  -0.0173  3.17  0.120   0.133    0.00799
##  5 Combina~ Trai~ -1.93e-2 0.518  0.413  -0.0805  3.21  0.122   0.136    0.0251
##  6 ARIMA2   Trai~ -1.75e-2 0.536  0.416  -0.138   3.23  0.122   0.140    0.0227
##  7 Combina~ Trai~ -2.59e-2 0.596  0.486  -0.428   3.69  0.143   0.156    0.429
##  8 Combina~ Trai~ -1.02e-2 0.598  0.484  -0.284   3.70  0.143   0.156    0.478
##  9 Combina~ Trai~ -8.74e-3 0.598  0.509  -0.359   3.94  0.150   0.156    0.426
## 10 Combina~ Trai~ -9.09e-3 0.605  0.508  -0.347   3.95  0.150   0.158    0.505
## # i 13 more rows
```

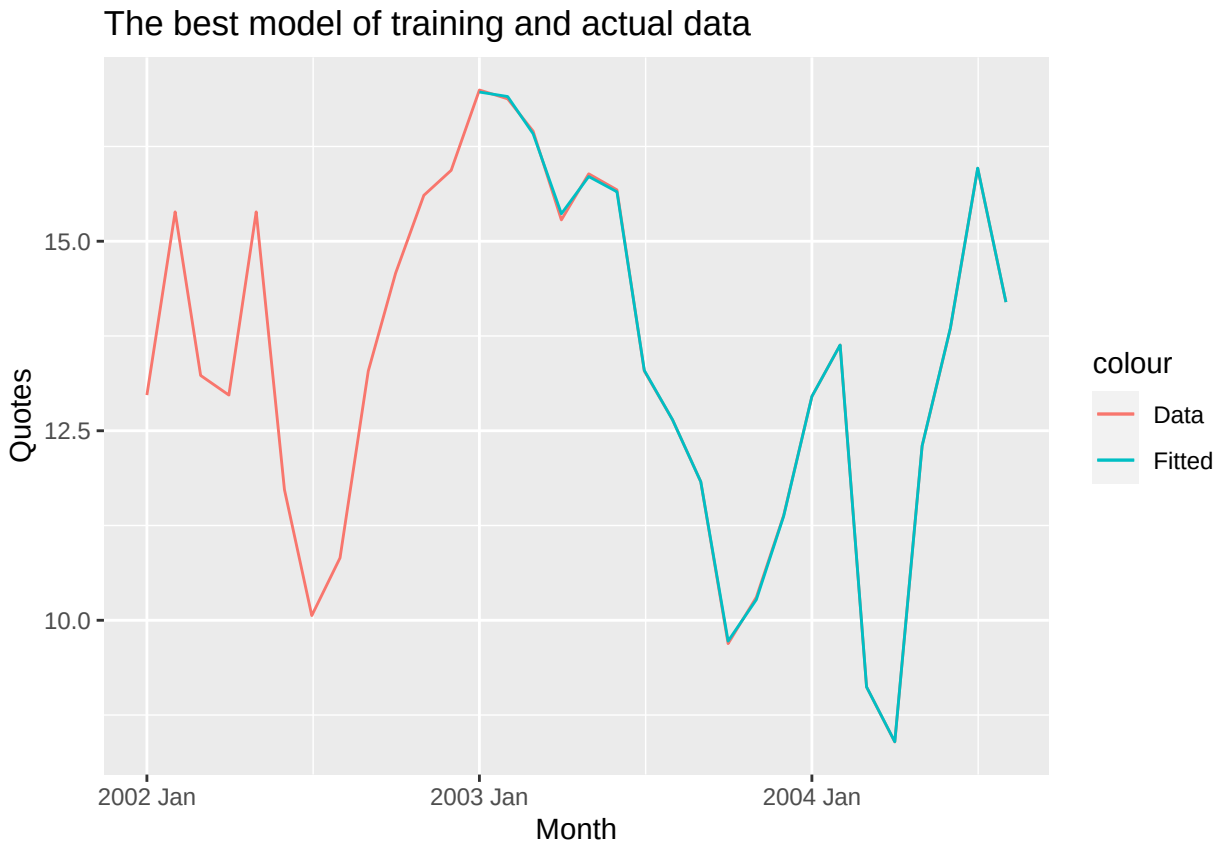
```
print(c("The best mdoel on Train is: ", acc2$.model[1]))
```

```
## [1] "The best mdoel on Train is: " "NNET"
```

```
fit2[,acc1$.model[1]] |> gg_tsresiduals()
```




```
augment(fit2) |> filter(.model == acc1$.model[1]) |>
  ggplot(aes(x = Month)) +
  geom_line(aes(y = Quotes , colour = "Data")) +
  geom_line(aes(y = .fitted, colour = "Fitted"))+
  ggtitle("The best model of training and actual data")
```



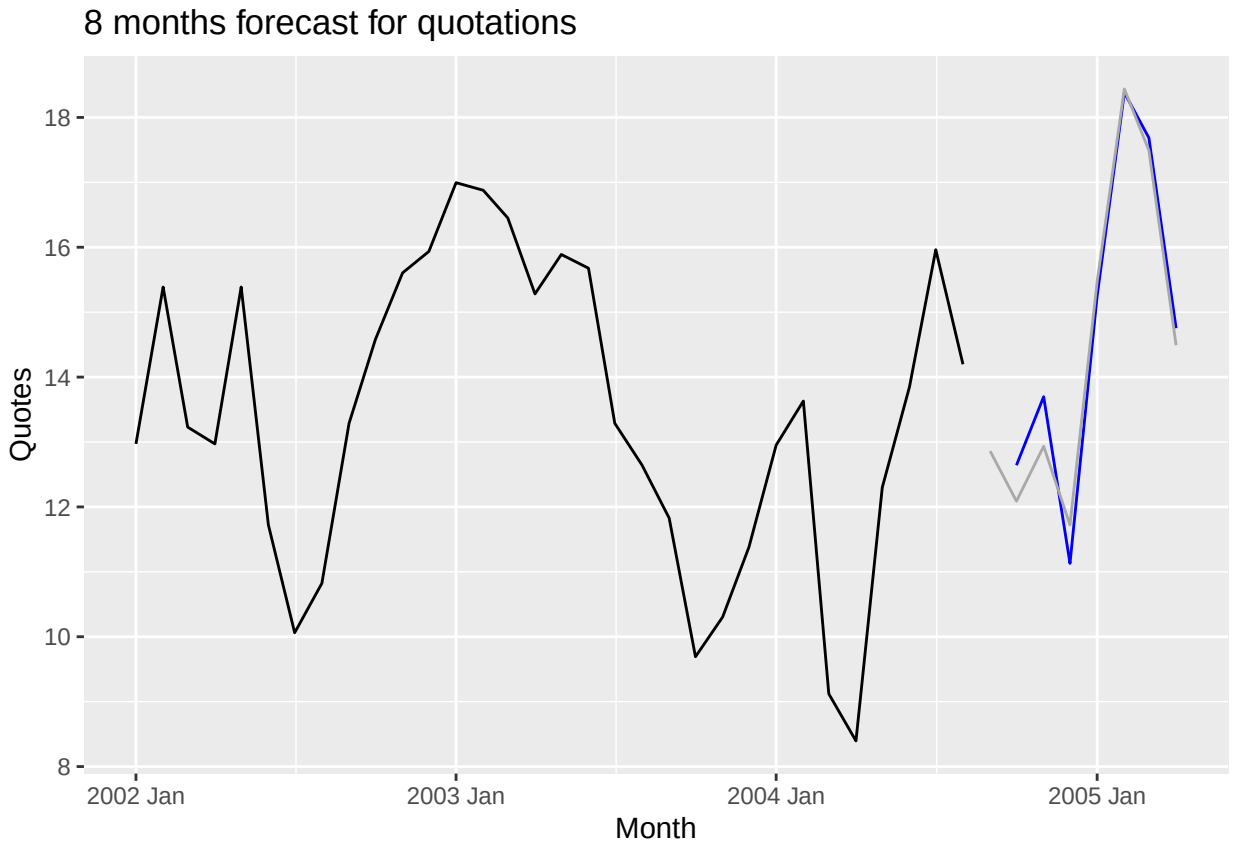

```
# augment(fit2) |> filter(.model == acc2$.model[1]) |>
#   features(.resid, ljung_box, lag=12, dof=0)
#
# augment(fit2) |> filter(.model == acc2$.model[1]) |>
#   features(.innov, box_pierce, dof = 0, lag = 12)


fc2 <- fit2 |> forecast(new_data =  Test)

accT2 <- accuracy(fc2, Test) |> arrange(RMSE)
accT2
```

```
## # A tibble: 23 x 10
##    .model       .type     ME  RMSE   MAE    MPE  MAPE  MASE RMSSE   ACF1
##    <chr>        <chr>  <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>  <dbl>
##  1 Combination1 Test  -0.129 0.447 0.377 -0.943  2.88   NaN   NaN 0.0983
##  2 lm2          Test  -0.163 0.465 0.352 -1.25   2.78   NaN   NaN 0.104
##  3 lm4          Test  -0.153 0.475 0.435 -1.10   3.23   NaN   NaN 0.240
##  4 Combination3 Test   0.244 0.564 0.519  1.37   3.55   NaN   NaN 0.570
##  5 Combination4 Test   0.203 0.577 0.521  0.989  3.53   NaN   NaN 0.502
##  6 Combination2 Test   0.144 0.582 0.547  0.573  3.76   NaN   NaN 0.609
##  7 Combination5 Test   0.299 0.584 0.507  1.77   3.43   NaN   NaN 0.457
##  8 Combination6 Test   0.551 0.926 0.741  3.20   4.72   NaN   NaN 0.681
##  9 ARIMA2       Test   0.451 0.935 0.769  2.40   4.93   NaN   NaN 0.693
## 10 ARIMA4       Test   0.650 0.943 0.713  4.00   4.51   NaN   NaN 0.657
## # i 13 more rows
```

```
print(c("The best mdoel on Test is: ", accT2$.model[1]))
```

```
## [1] "The best mdoel on Test is: " "Combination1"
```

```
fc2 |> filter(.model == accT2$.model[1]) |>
  autoplot(Train, level = NULL) +
  labs(y = "Quotes", x = "Month",
  title = "8 months forecast for quotations") +
    autolayer(Test, colour = "darkgray")
```



Cross Validation

```
insurance_stretch <- insurance |>
  stretch_tsibble(.init = 5, .step = 1) |>
  filter(.id != max(.id))


fit_cv1 <- insurance_stretch |>
   model(
        lm     =  TSLM(Quotes),
        lm2    =  TSLM(Quotes~ TVadverts),
        lm3    =  TSLM(Quotes~ lag(TVadverts)),
        lm4    =  TSLM(Quotes~ TVadverts + lag(TVadverts)),
        Mean   =  MEAN(Quotes),
        SNaive =  SNAIVE(Quotes),
        Naive  =  NAIVE(Quotes),
        Drift  =  RW(Quotes ~ drift()),
```

```
        ETS    =  ETS(Quotes),
        NNET   =  NNETAR(Quotes),
        Prophet = prophet(Quotes)
)

fc_cv1 <- fit_cv1 |>
  forecast(new_data = insurance_stretch , h=8)


# Cross-validated
fc_cv1 |> accuracy(insurance) |> arrange(RMSE)
```

```
## # A tibble: 11 x 10
##    .model  .type        ME  RMSE   MAE    MPE  MAPE  MASE RMSSE  ACF1
##    <chr>   <chr>     <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 Prophet Test  -1.92e- 5 0.752 0.383 -0.365  2.96   NaN   NaN 0.317
##  2 lm4     Test  -3.74e-16 0.813 0.684 -0.406  5.19   NaN   NaN 0.743
##  3 lm2     Test  -3.13e-16 0.833 0.714 -0.431  5.40   NaN   NaN 0.564
##  4 lm3     Test   4.20e-16 1.97  1.68  -2.39  13.2    NaN   NaN 0.372
##  5 Mean    Test   1.31e-16 2.19  1.83  -2.92  14.4    NaN   NaN 0.643
##  6 lm      Test   1.48e-16 2.19  1.83  -2.92  14.4    NaN   NaN 0.643
##  7 SNaive  Test   4.31e- 1 3.20  2.53   0.481 19.2    NaN   NaN 0.595
##  8 ETS     Test   1.63e- 2 3.32  2.63  -2.79  20.3    NaN   NaN 0.844
##  9 Naive   Test   1.53e- 3 3.36  2.68  -2.91  20.7    NaN   NaN 0.849
## 10 NNET    Test   1.04e+ 0 4.15  3.10   5.10  23.0    NaN   NaN 0.755
## 11 Drift   Test  -3.32e- 1 4.56  3.69  -5.30  28.4    NaN   NaN 0.920
```

```
        # ARIMA  =  ARIMA(Quotes),
        # ARIMA2 =  ARIMA(Quotes~ TVadverts),
        # ARIMA3 =  ARIMA(Quotes~ lag(TVadverts)),
        # ARIMA4 =  ARIMA(Quotes~ TVadverts + lag(TVadverts)),
```

Bootstapped

```
sim <- ins_stl |>
  generate(new_data = Train, times = 100,
           bootstrap_block_size = 24) |>  select(-.model, -Quotes)


fit3 <-  sim |>  model(
        lm     =  TSLM(.sim),
        lm2    =  TSLM(.sim ~ TVadverts),
        lm3    =  TSLM(.sim ~ lag(TVadverts)),
        lm4    =  TSLM(.sim ~ TVadverts + lag(TVadverts)),
        ARIMA  =  ARIMA(.sim),
        ARIMA2 =  ARIMA(.sim ~ TVadverts),
        ARIMA3 =  ARIMA(.sim ~ lag(TVadverts)),
        ARIMA4 =  ARIMA(.sim ~ TVadverts + lag(TVadverts)),
        Mean   =  MEAN(.sim),
        SNaive =  SNAIVE(.sim),
        Naive  =  NAIVE(.sim),
        Drift  =  RW(.sim ~ drift()),
        ETS    =  ETS(.sim),
```

```
        NNET   =  NNETAR(.sim),
        Prophet = prophet(.sim)
)

acc3 <- accuracy(fit3) |> arrange(RMSE)

acc3
```

```
## # A tibble: 1,500 x 11
##    .rep  .model  .type          ME     RMSE      MAE       MPE     MAPE     MASE
##    <chr> <chr>   <chr>       <dbl>    <dbl>    <dbl>     <dbl>    <dbl>    <dbl>
##  1 40    Prophet Traini~  3.11e-15 3.11e-13 2.45e-13 -3.71e-14 1.85e-12 1.07e-13
##  2 50    Prophet Traini~ -6.93e-14 3.35e-13 2.58e-13 -5.74e-13 1.99e-12 9.52e-14
##  3 82    Prophet Traini~ -4.00e-15 9.80e-13 7.94e-13 -6.23e-14 6.48e-12 2.71e-13
##  4 58    Prophet Traini~  2.93e-14 1.09e-12 8.95e-13  6.09e-13 6.91e-12 3.74e-13
##  5 36    Prophet Traini~ -7.22e-13 3.90e-12 3.21e-12 -4.97e-12 2.43e-11 1.44e-12
##  6 14    Prophet Traini~ -1.05e-12 4.42e-12 3.46e-12 -7.57e-12 2.68e-11 1.07e-12
##  7 8     NNET    Traini~ -1.99e- 5 6.64e- 4 5.57e- 4 -2.00e- 4 4.45e- 3 3.29e- 4
##  8 47    NNET    Traini~  9.75e- 7 9.70e- 4 7.70e- 4 -5.41e- 5 5.78e- 3 2.23e- 4
##  9 55    NNET    Traini~  3.01e- 3 2.73e- 2 2.09e- 2  1.74e- 2 1.55e- 1 1.42e- 2
## 10 25    NNET    Traini~  7.61e- 5 3.08e- 2 1.28e- 2 -1.38e- 3 8.71e- 2 3.23e- 3
## # i 1,490 more rows
## # i 2 more variables: RMSSE <dbl>, ACF1 <dbl>
```

```
fc3 <- fit3 |>filter(.rep == acc3$.rep[1]) |> select(-.rep) |> forecast(new_data = Test)

names(Test)[names(Test) == "Quotes"] <- ".sim"

accT3 <- accuracy(fc3, Test) |> arrange(RMSE)
accT3
```

```
## # A tibble: 15 x 10
##    .model  .type     ME  RMSE   MAE   MPE  MAPE  MASE RMSSE   ACF1
##    <chr>   <chr>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>
##  1 lm2     Test   0.695  1.61  1.28  3.38  8.18   NaN   NaN  0.632
##  2 lm4     Test   0.810  1.68  1.37  4.04  8.68   NaN   NaN  0.503
##  3 ARIMA2  Test   0.592  2.15  1.94  2.27 13.3    NaN   NaN  0.589
##  4 ARIMA4  Test   0.641  2.20  1.97  2.58 13.5    NaN   NaN  0.598
##  5 SNaive  Test   1.04   2.22  1.45  5.71  9.06   NaN   NaN  0.524
##  6 lm3     Test   1.11   2.48  1.96  5.32 12.5    NaN   NaN  0.379
##  7 Mean    Test   1.18   2.63  2.04  5.84 12.9    NaN   NaN  0.568
##  8 lm      Test   1.18   2.63  2.04  5.84 12.9    NaN   NaN  0.568
##  9 ARIMA3  Test   2.67   3.10  2.67 18.0  18.0    NaN   NaN  0.431
## 10 ETS     Test   4.00   4.64  4.00 25.9  25.9    NaN   NaN  0.568
## 11 Naive   Test   4.00   4.64  4.00 25.9  25.9    NaN   NaN  0.568
## 12 Drift   Test   4.38   5.04  4.38 28.4  28.4    NaN   NaN  0.600
## 13 NNET    Test  -0.634  5.36  4.41 -9.33 32.1    NaN   NaN  0.550
## 14 Prophet Test  -0.286  6.39  4.28 -5.92 32.6    NaN   NaN -0.170
## 15 ARIMA   Test   6.49   6.83  6.49 44.3  44.3    NaN   NaN  0.537
```

```
print(c("The best mdoel on Test is: ", accT3$.model[1]))
```

```
## [1] "The best mdoel on Test is: " "lm2"
```