Signal extraction examples

Sonia Mazzi

13/12/2018

Signal extraction using a state-space model

Suppose that we have $y = \{y_t\}_{t=1}^{n}$ and $y_t = \mu_t + \epsilon_t$, where μ_t is a stochastic process (with linear dynamic) and ϵ_t's are iid with mean zero and constant variance. Note that y_t can be a vector.

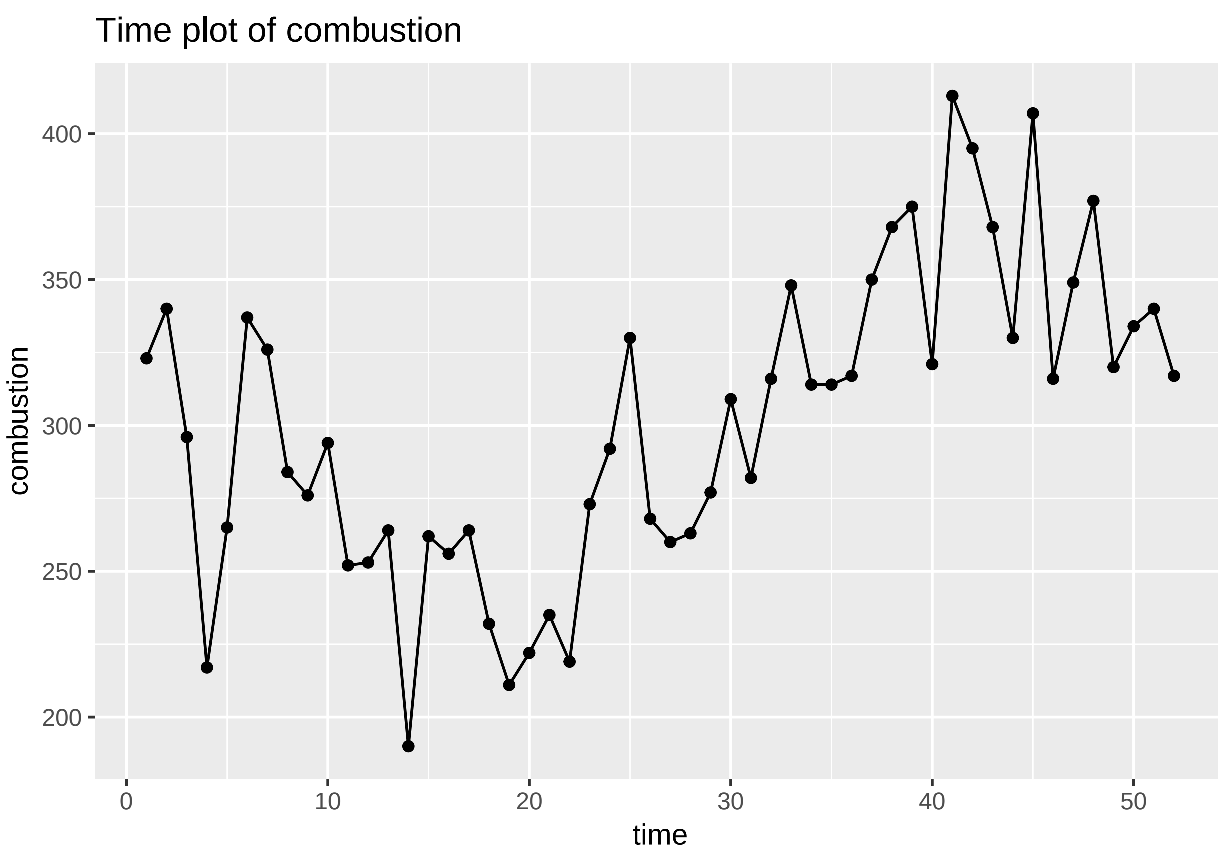
The stochastic specification of μ_t allows to model y in different ways.

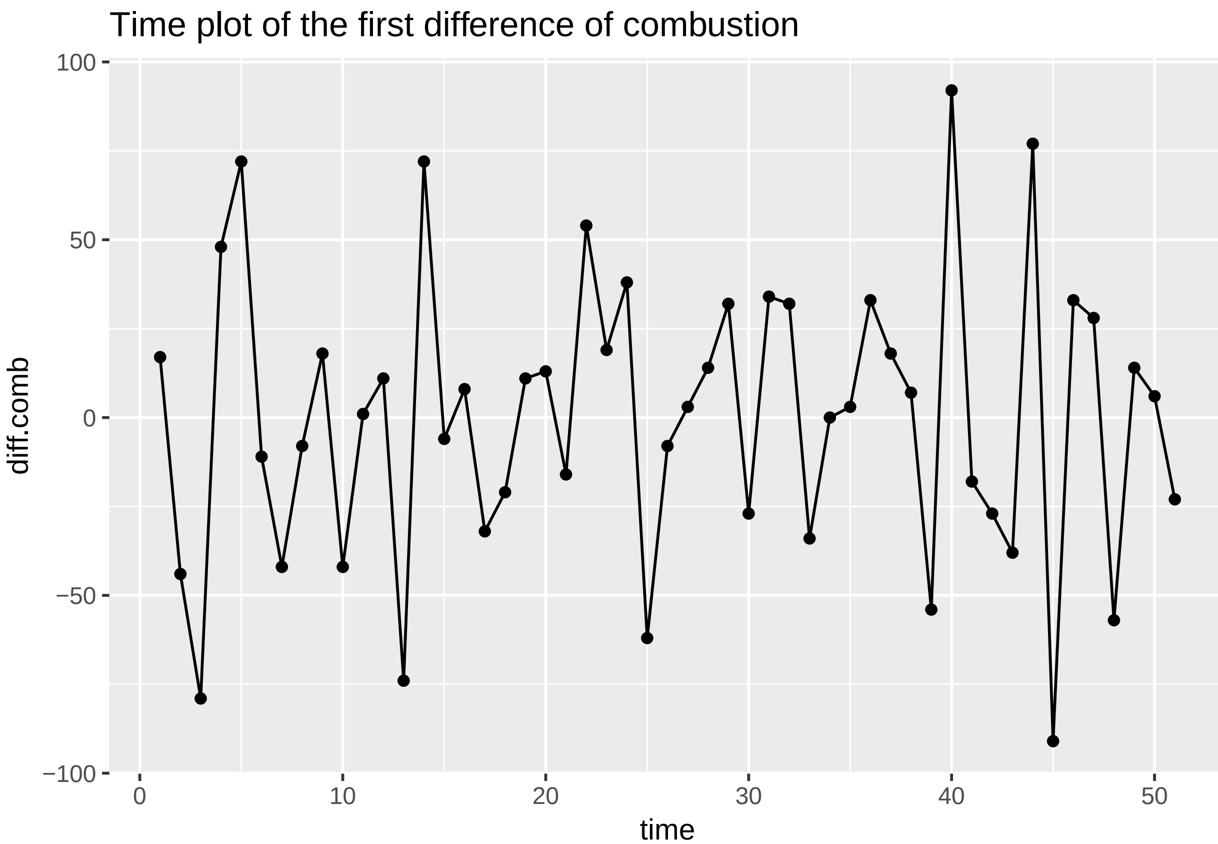
For more details on the state-space model and signal-extraction algorithms see

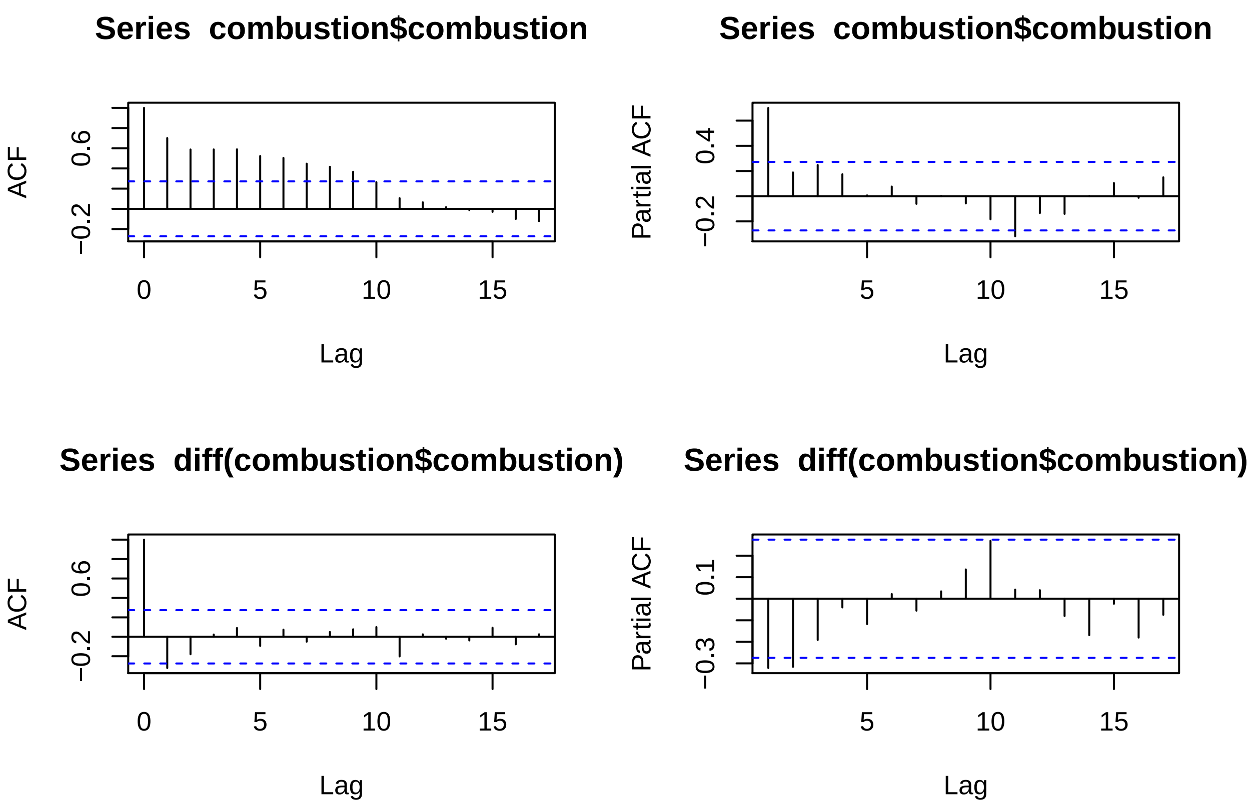
de Jong, P.(1991). The diffuse Kalman filter. *The Annals of Statistics.* Vol 19, No.2, pp 1073 - 1083.

I will illustrate the signal extraction prodedure with three sample quarterly time series: `combustion`, and `image`.

`combustion`







We can observe that the plots above indicate that `combustion` is a 'pure trend' series, with some extreme observations (possibly caused by other processes).

First I will fit a general locally linear model (LLM), which is essentially a random walk with a time varying drift.

In the LLM it is assumed

$y_t = \mu_t + \epsilon_t$, $\mu_{t+1} = \mu_t + d_t + \nu_t$, $d_{t+1} = \delta d_t + \eta_t$

where $\{\epsilon_t\}$, $\{\nu_t\}$, $\{\eta_t\}$ are mutually independent white noise processes with variances σ^2_ϵ, σ^2_ν and σ^2_η.

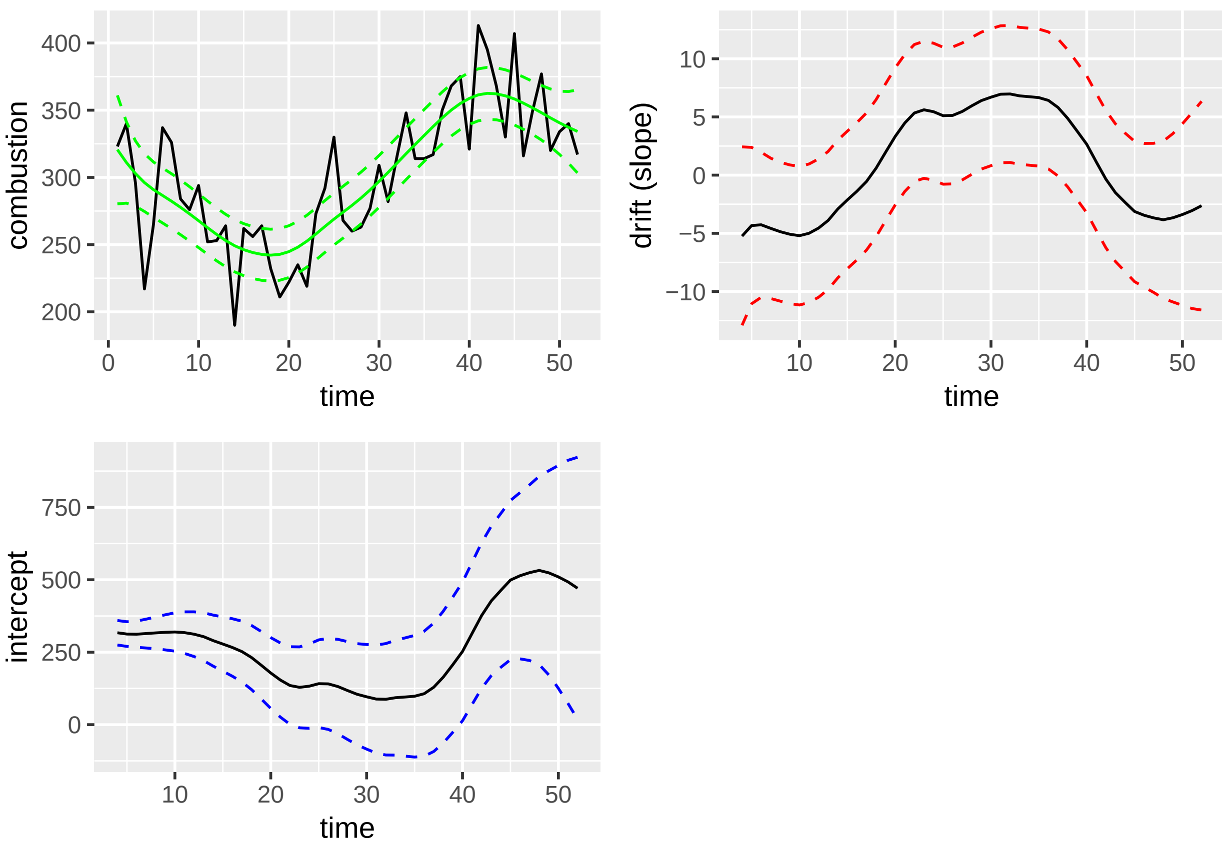
Fitting the LLM for `combustion`

Estimated parameters

Parameter	Estimate
σ_ϵ	29.7853952
σ_ν	0.0000019
σ_η	2.6454416
δ	0.8608259

Plots of predicted signal

```r
grid.arrange(p1, p2, p3, nrow = 2)
```

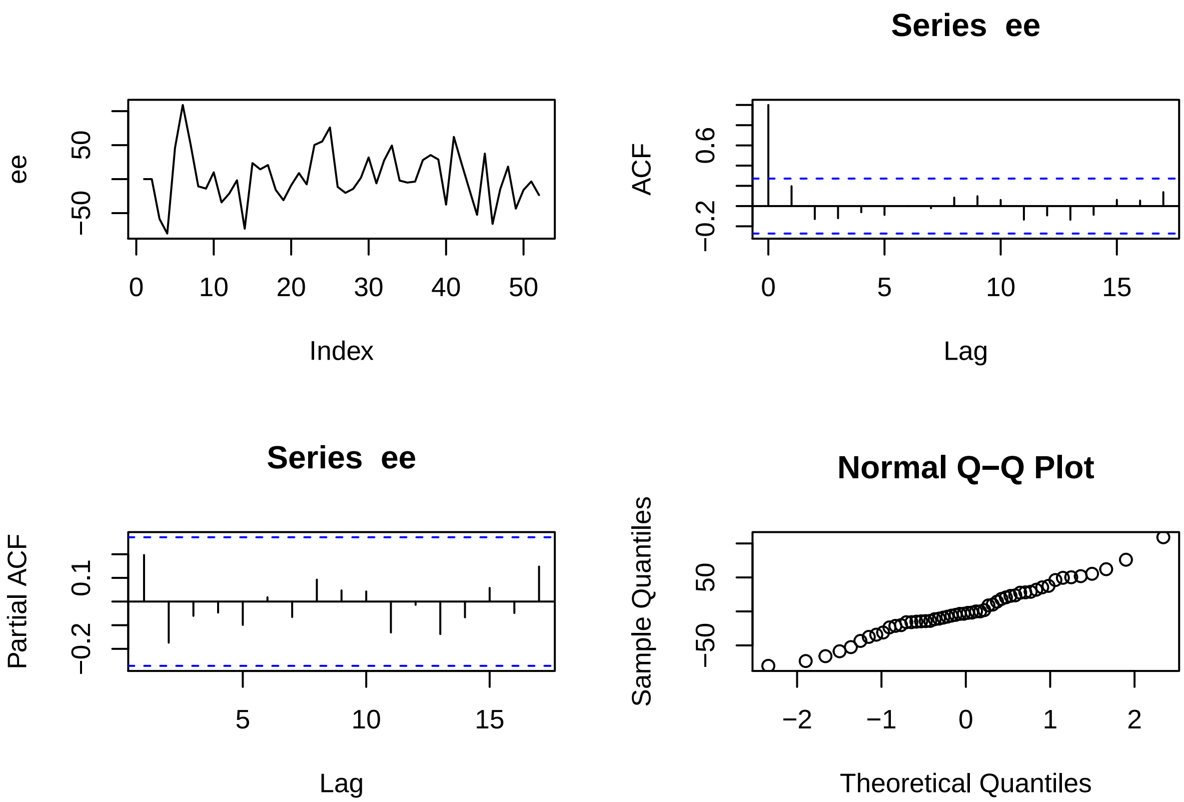


In the above plots, the green solid line has predicted μ_ts. All dotted lines are approximate 95% confidence bands (predicted value$\pm 2\sqrt{mse}$).

The "drift" is d_t and the "intercept" is $\mu_t - td_t$. The drift can be interpreted as the first derivative of the ideal smooth trend. So, when the drift is positive there is growth and when the drift is negative the trend is decreasing.

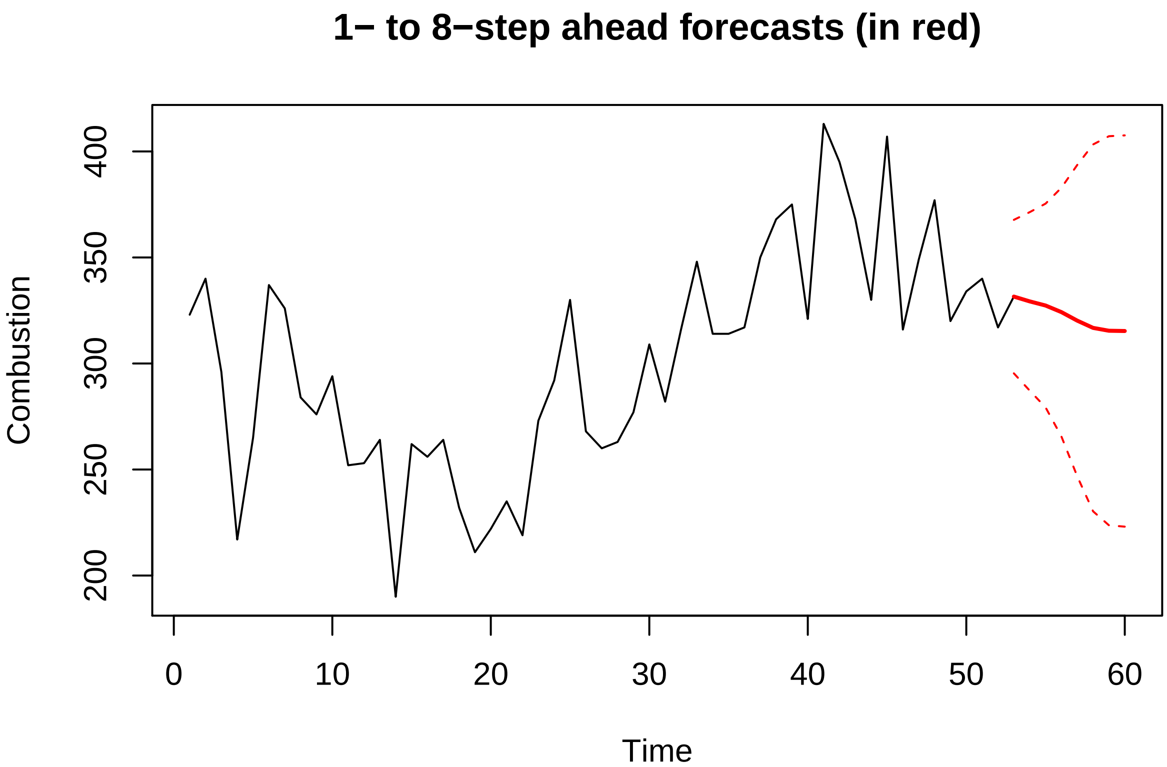
Note that for `combustion` the last few quarters, although the drift is negative, it has started to increase, possibly announcing that the decay in trend is decelarting and a period of growth in trend is coming once the drift crosses the zero line. This is re-inforced by an accompanying decrease in "intercept".

Residual Analysis



The plots above don't indicate that the model isn't fitting the data well.

Forecasting



Prediction accuracy

We fit the model for the series except the last four observations and predict the next four observations.

Let $\hat{y}_{n+i}$ be the predictor of y_{n+i} using $y_1, \ldots, y_n$, $i = 1, \ldots, k$.

If $k = 4$, then

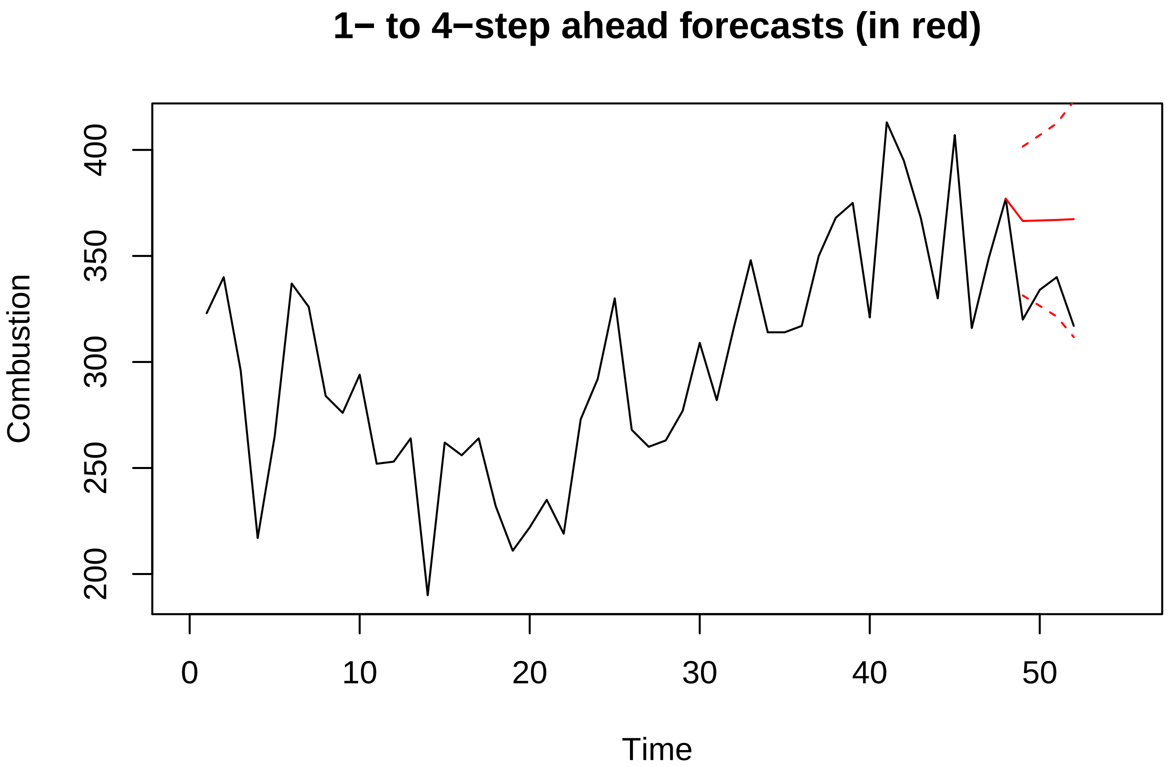
i	y_{n+i}	$\hat{y}_{n+i}$
0	377	377.0000
1	320	366.4679
2	334	366.7288
3	340	366.9636
4	317	367.3652

We define $e_{n+i} = y_{n+i} - \hat{y}_{n+i}$, $i = 1, \ldots, k$, the prediction errors. Then, the mean squared error of prediction is $\frac{1}{n}\sum_{i=1}^{k}(e_{n+1})^2$ and the mean absolute percentage error is $\frac{1}{n}\sum_{i=1}^{k} 100|e_{n+i}|/|y_{n+i}|$.

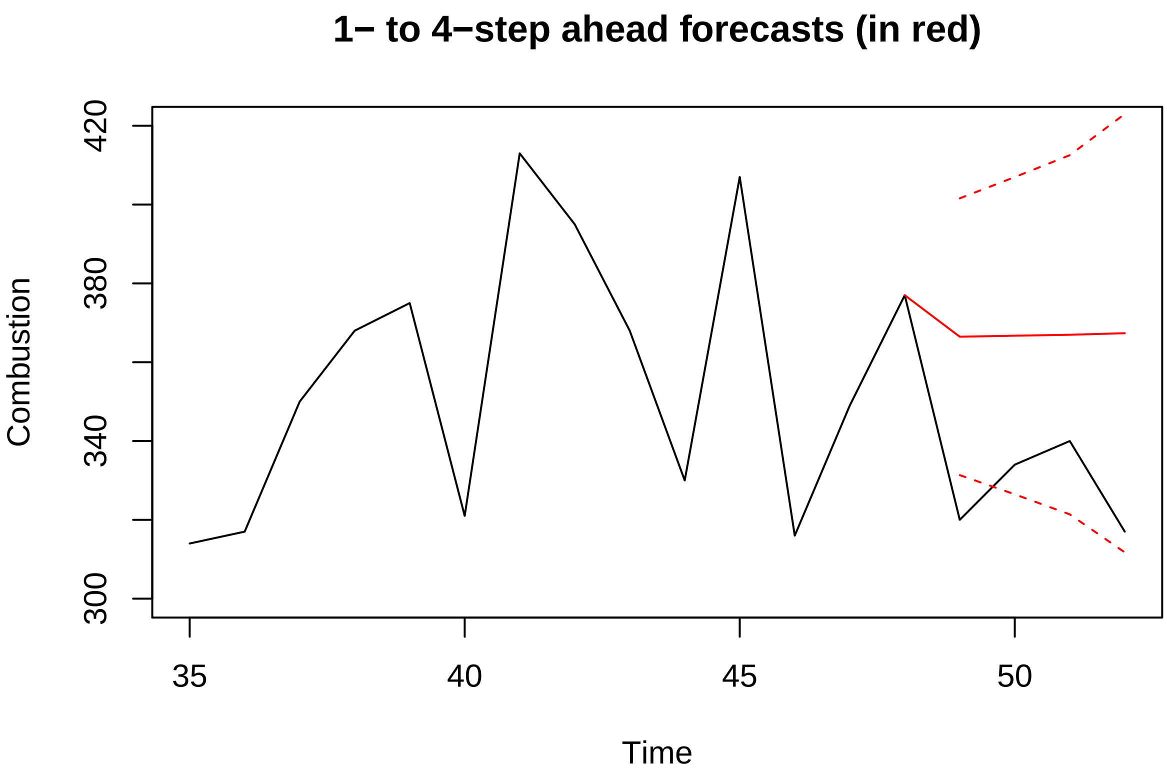
From the data:

The square root of the Mean Squared Error is 40.2930675.

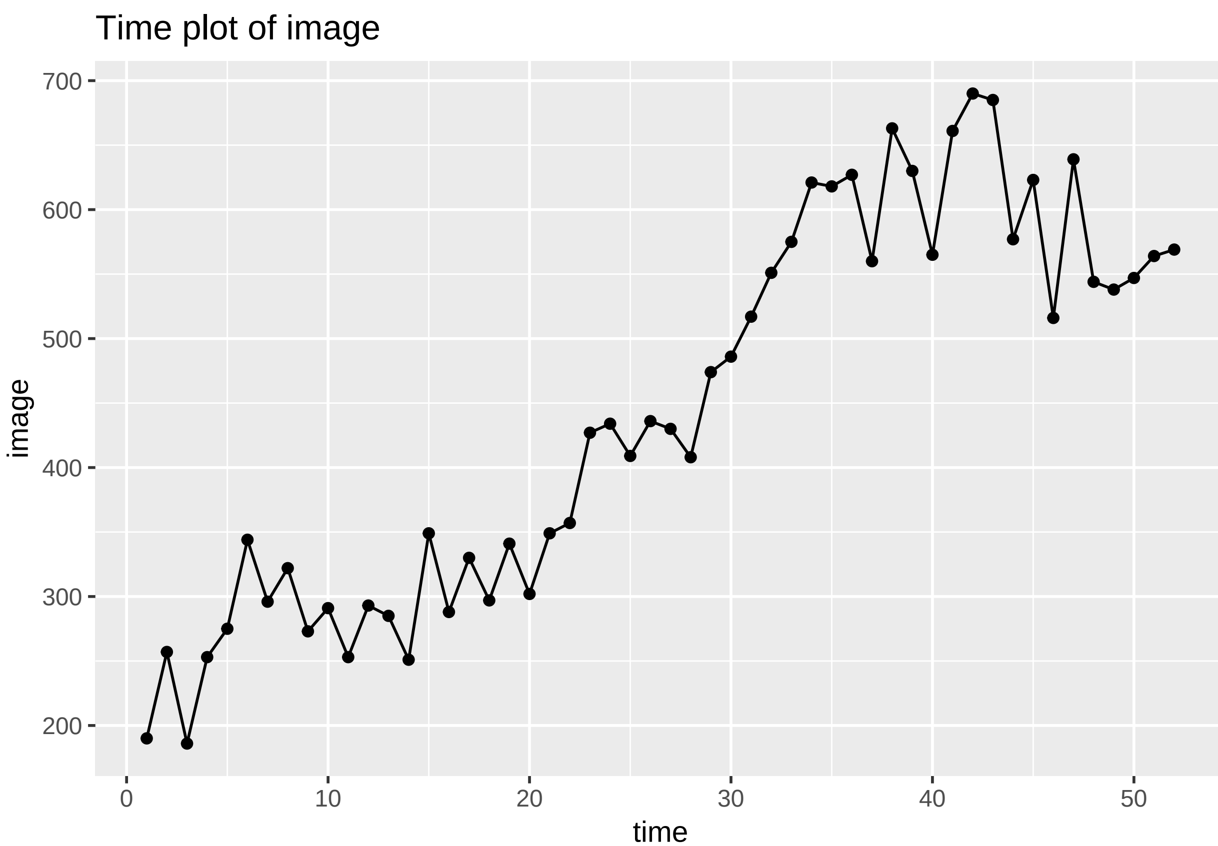
The mean Absolute Percentage Error is 12.0346971.

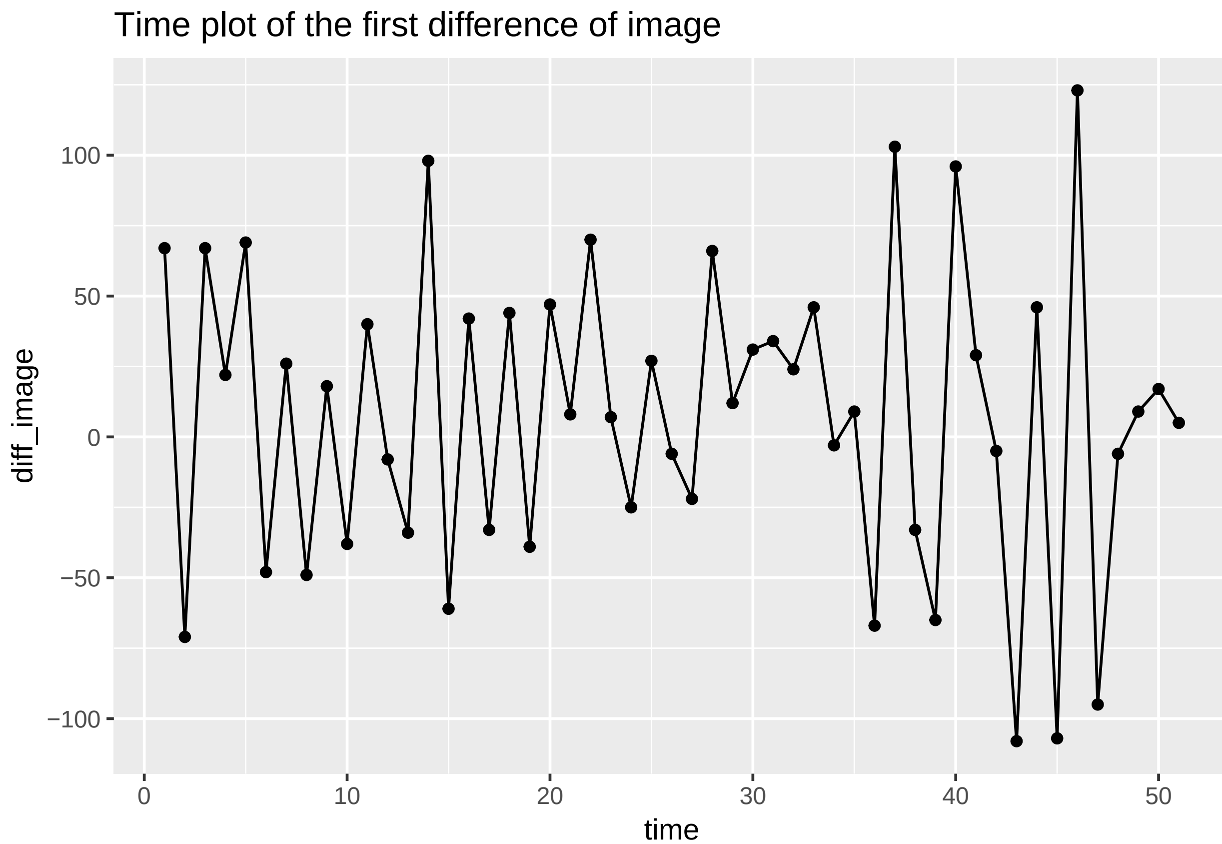


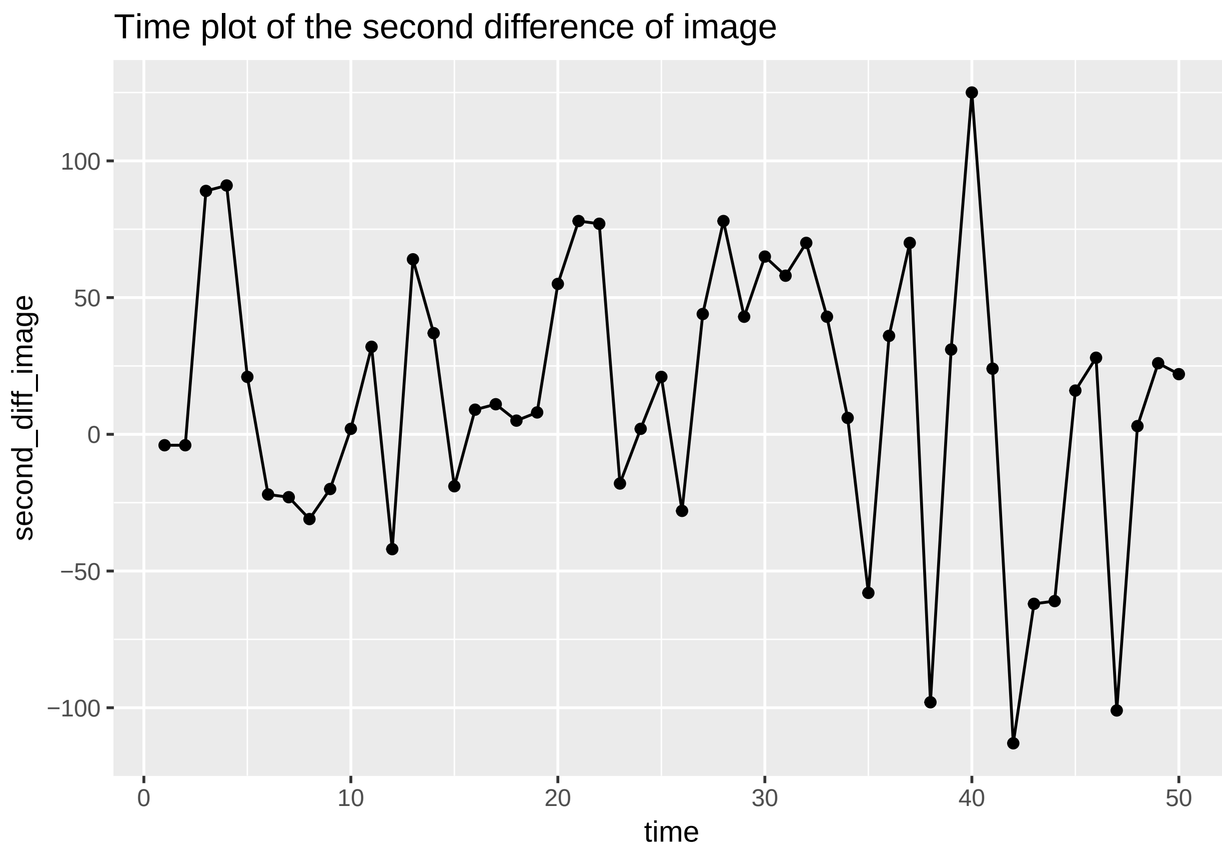
Let us zoom-in

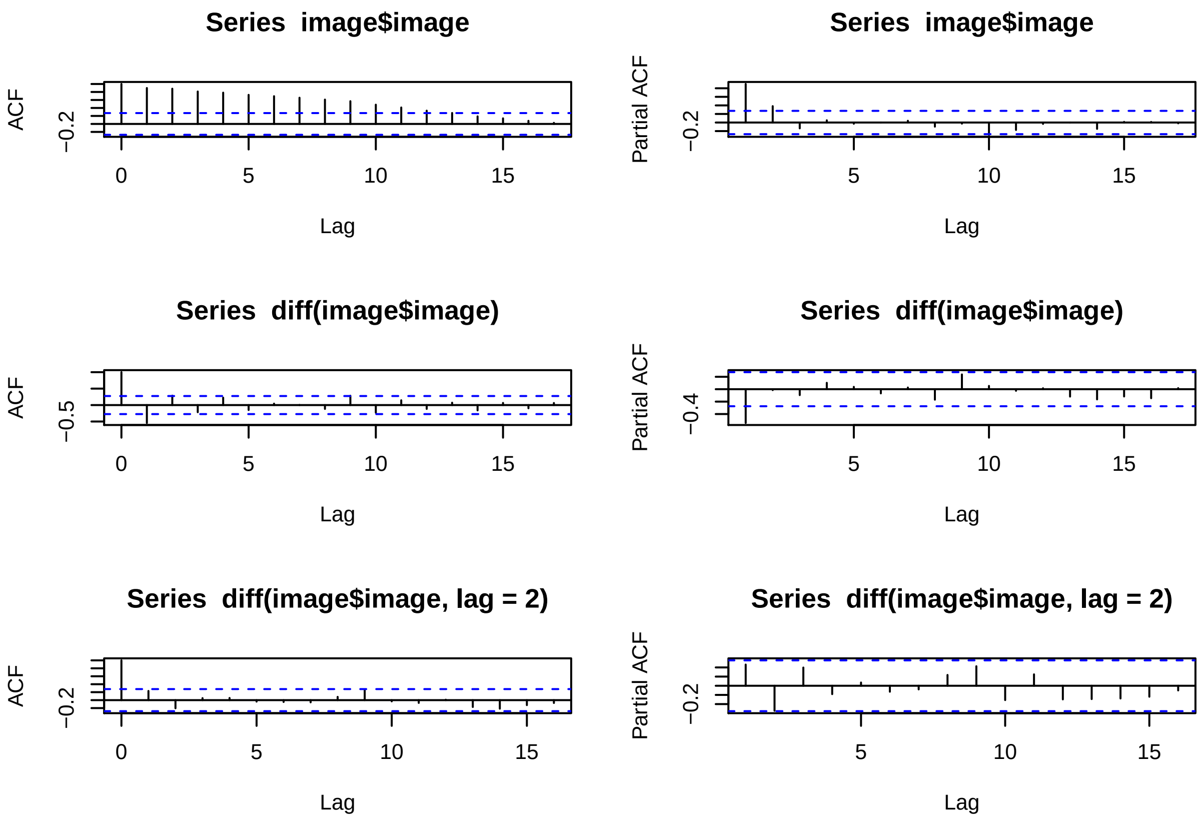


`image`


Time plot of image
image
700
600
500
400
300
200
0
10
20
30
40
50
time


Time plot of the first difference of image
diff_image
100
50
0
−50
−100
0
10
20
30
40
50
time


Time plot of the second difference of image
second_diff_image
100
50
0
−50
−100
0
10
20
30
40
50
time



We can observe that the plots above indicate that the autocorrelation function of the first difference is significant at lag one. Also, from the time plot of the first difference we clearly see that the series has a change in variance (it is heteroscedastic). So, a LLM would not be a appropriate. Instead a locally quadratic model (LQM) will be fitted.

In the LQM it is assumed

$y_t = \mu_t + \epsilon_t$, $\mu_{t+2} = 2\mu_{t+1} - \mu_t + d_t + \nu_t$, $d_{t+1} = \delta d_t + \eta_t$

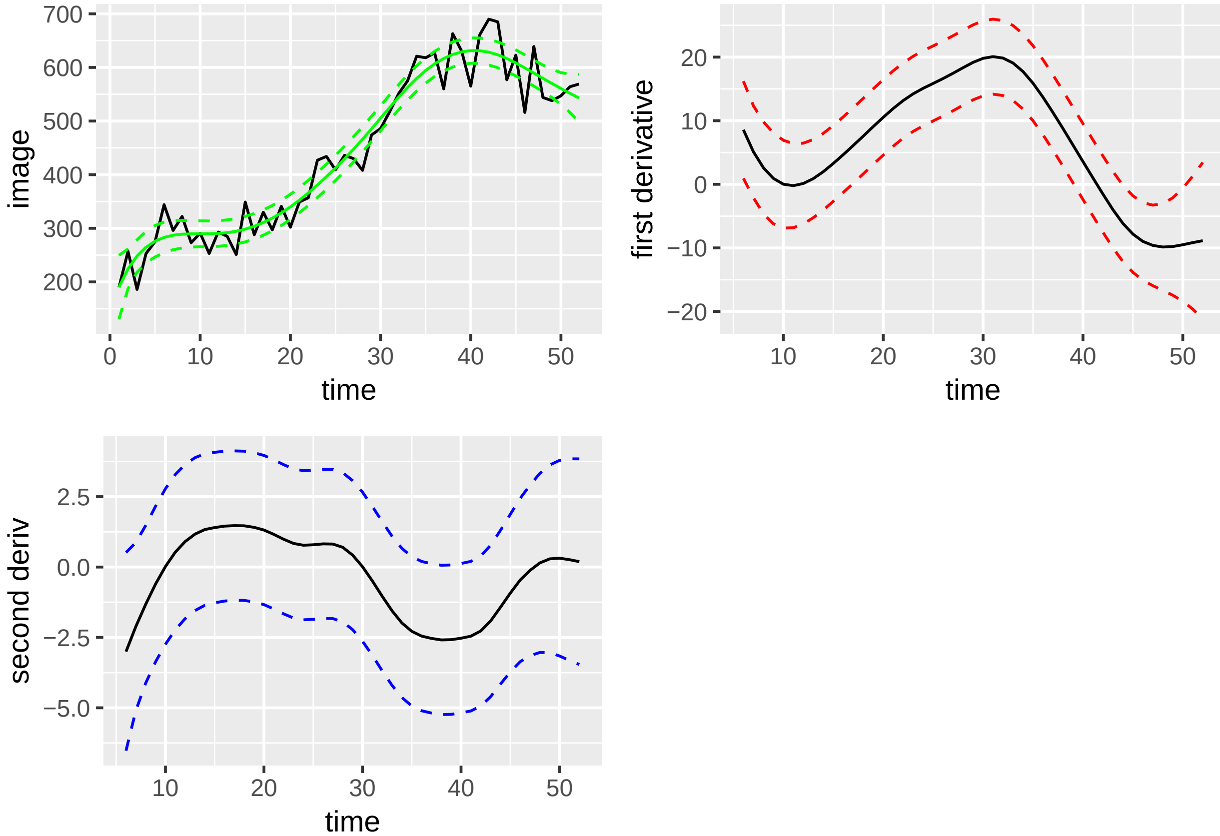
where $\{\epsilon_t\}$, $\{\nu_t\}$, $\{\eta_t\}$ are mutually independent white noise processes with variances σ^2_ϵ, σ^2_ν and σ^2_η.

Fitting the LQM for `image`

Estimated parameters

Parameter	Estimate
σ_ϵ	36.4279645
σ_ν	0.0000007
σ_η	1.2718623
δ	0.7311585

Plots of predicted signal

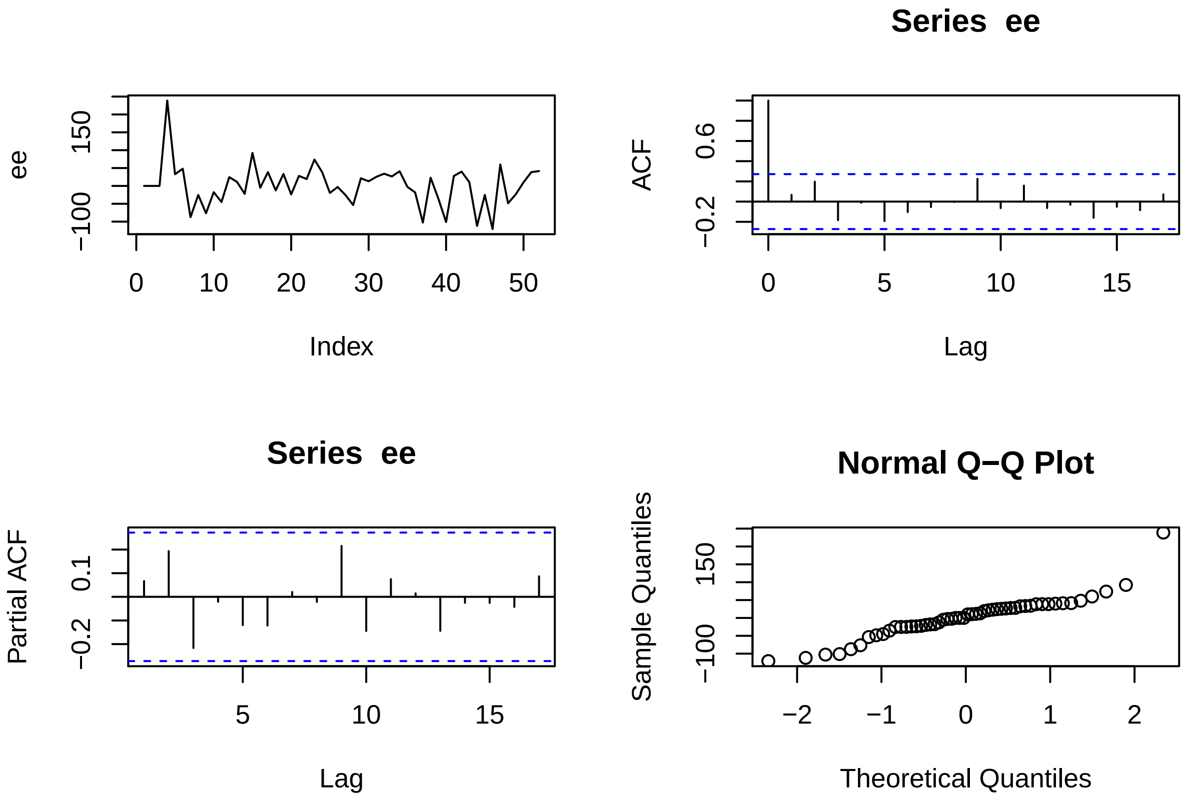


In the above plots, the green solid line has predicted μ_ts. All dotted lines are approximate 95% confidence bands (predicted value$\pm 2\sqrt{mse}$).

The "first derivative" is $\mu_t - \mu_{t-1} - 0.5 d_t$ and the "second derivative" is d_t.

Note that for `image` the last few quarters, although the drift is negative, it has started to increase, possibly announcing that the downfall trend is decelerating and may be a period of growth in trend is coming.

Residual Analysis



The plots above don't indicate that the model isn't fitting the data well.

Forecasting

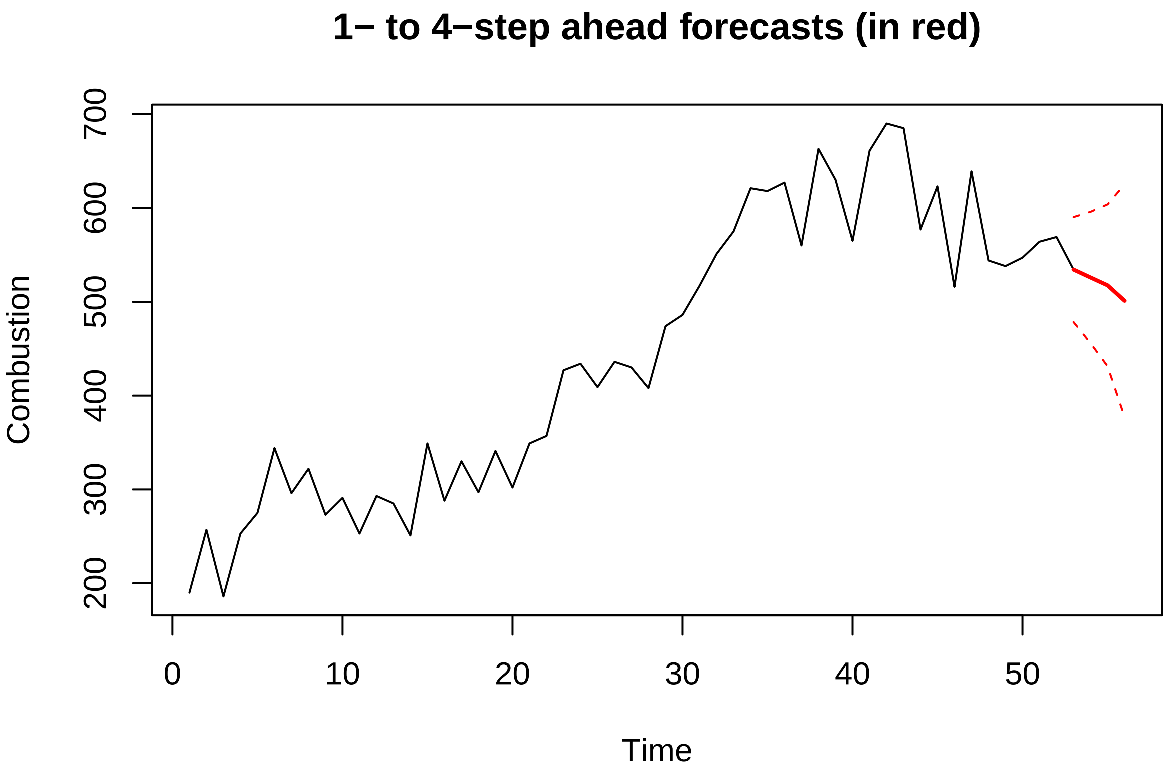
```
source("forecast")
k=4
forecast.out.image = forecast(dkf.out.image, k)
```

```
yy = image$image
lyy = length(yy)
yy_f = c(image$image, forecast.out.image$yhat)
yy_f
```

```
##  [1] 190.0000 257.0000 186.0000 253.0000 275.0000 344.0000 296.0000
##  [8] 322.0000 273.0000 291.0000 253.0000 293.0000 285.0000 251.0000
## [15] 349.0000 288.0000 330.0000 297.0000 341.0000 302.0000 349.0000
## [22] 357.0000 427.0000 434.0000 409.0000 436.0000 430.0000 408.0000
## [29] 474.0000 486.0000 517.0000 551.0000 575.0000 621.0000 618.0000
## [36] 627.0000 560.0000 663.0000 630.0000 565.0000 661.0000 690.0000
## [43] 685.0000 577.0000 623.0000 516.0000 639.0000 544.0000 538.0000
## [50] 547.0000 564.0000 569.0000 534.3116 525.8697 517.5302 501.0556
```

```
hi = forecast.out.image$yhat - 2 * sqrt(forecast.out.image$mse_yhat)
lo = forecast.out.image$yhat + 2 * sqrt(forecast.out.image$mse_yhat)
plot(yy_f, type="l", ylab = "Combustion", xlab = "Time", main = "1- to 4-step ahead forecasts (in red)"
lines(seq(lyy+1, lyy+k, 1), hi, lty = 2, col= "red")
```

```r
lines(seq(lyy+1, lyy+k, 1), lo, lty = 2, col= "red")
lines(seq(lyy+1, lyy+k, 1), forecast.out.image$yhat, lwd = 2, col= "red")
```




```r
image$image
```

```
##  [1] 190 257 186 253 275 344 296 322 273 291 253 293 285 251 349 288 330
## [18] 297 341 302 349 357 427 434 409 436 430 408 474 486 517 551 575 621
## [35] 618 627 560 663 630 565 661 690 685 577 623 516 639 544 538 547 564
## [52] 569
```

Prediction accuracy

We fit the model for the series except the last four observations and predict the next four observations.

Let $\hat{y}_{n+i}$ be the predictor of y_{n+i} using $y_1, \ldots, y_n$, $i = 1, \ldots, k$.

If $k = 4$, then

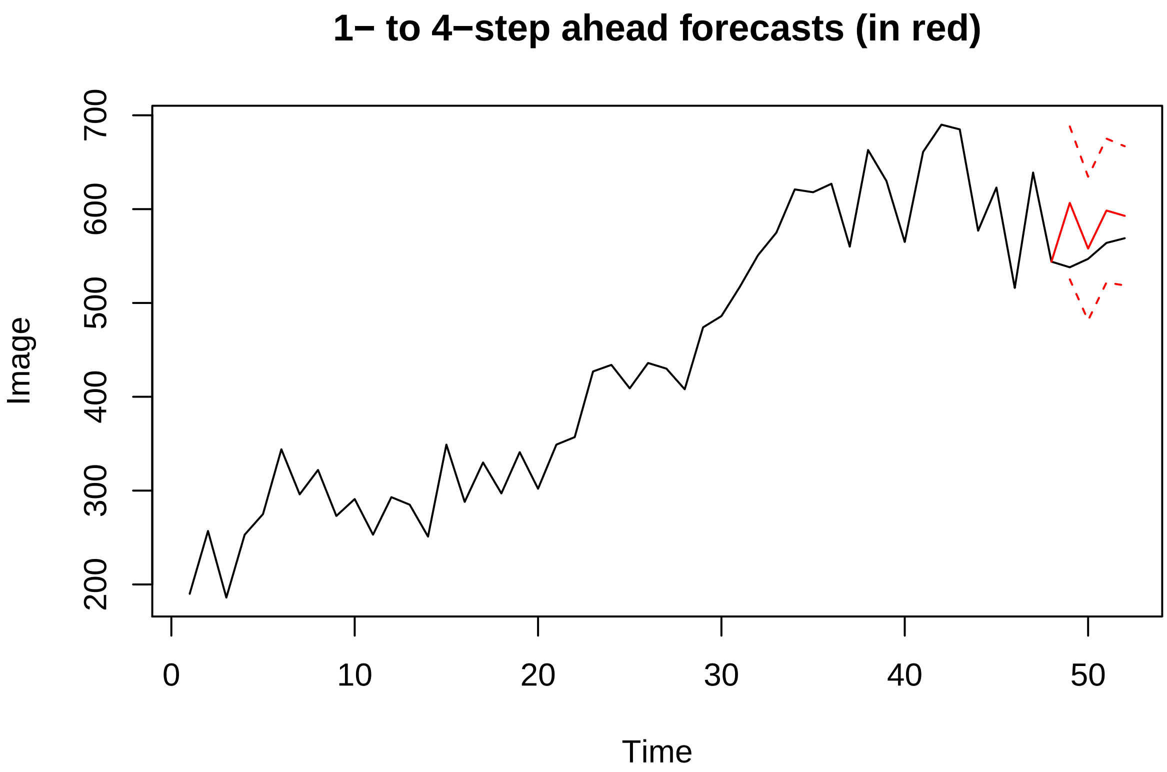
i	y_{n+i}	$\hat{y}_{n+i}$
0	544	544.0000
1	538	606.6393
2	547	557.9789
3	564	598.4861
4	569	592.8362

We define $e_{n+i} = y_{n+i} - \hat{y}_{n+i}$, $i = 1, \ldots, k$, the prediction errors. Then, the mean squared error of prediction is $\frac{1}{n}\sum_{i=1}^{k}(e_{n+1})^2$ and the mean absolute percentage error is $\frac{1}{n}\sum_{i=1}^{k} 100|e_{n+i}|/|y_{n+i}|$.

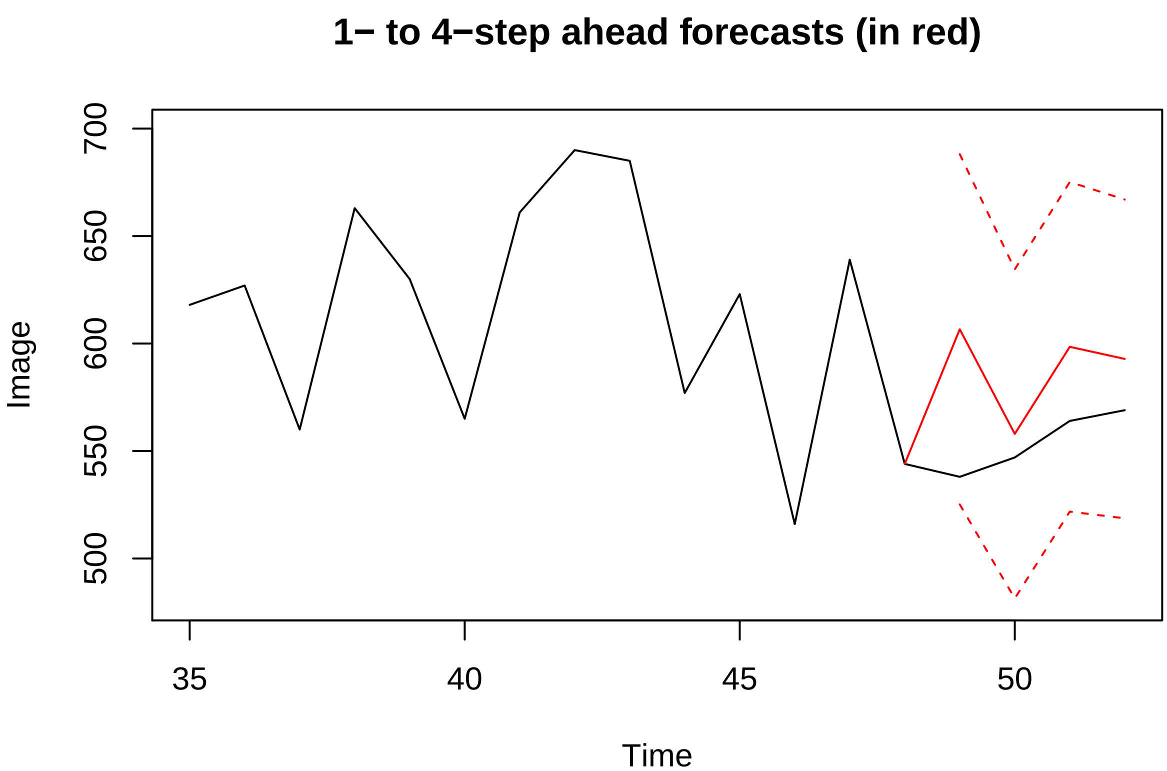
From the data:

The square root of the Mean Squared Error is 40.5873984.

The mean Absolute Percentage Error is 6.2672608.



Let us zoom-in




```
image$image
```

```
##  [1] 190 257 186 253 275 344 296 322 273 291 253 293 285 251 349 288 330
```

```
## [18] 297 341 302 349 357 427 434 409 436 430 408 474 486 517 551 575 621
## [35] 618 627 560 663 630 565 661 690 685 577 623 516 639 544 538 547 564
## [52] 569
```