Autoregressive model - AR(p)

Son Nguyen

AR(p)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon$$

β_i must satisfy certain conditions so that Y_t is stationary.

Partial Autocorrelation (pacf)

If X, Y and Z are random variables then the partial autocorrelation between X and Y given Z is the correlation between X and Y with the linear effect of Z removed from both X and Y.

- Regress X on Z to obtain $\hat{X}$, the linear effect of Z in X,
- $X - \hat{X}$ is X with the linear effect of Z removed
- Regress Y on Z to obtain $\hat{Y}$, the linear effect of Z in Y,
- $Y - \hat{Y}$ is Y with the linear effect of Z removed

(X, Y) linear relation

with Z | without Z

$$p_{XY|Z} = corr(X - \hat{X}, Y - \hat{Y})$$

(X Z) there may be some linear relation

(Y Z) there may be some linear relation

Remove both of these linear relation then calculate the linear relation between (X, Y)

① Regress X on Z

$$X \approx \hat{X} = \beta_0 + \beta_1 Z \Rightarrow X - \hat{X} \quad \text{residual } 1$$

② Regress Y on Z

$$Y \approx \hat{Y} = \beta_0^* + \beta_1^* Z$$

$$\Rightarrow Y - \hat{Y} \quad \text{residual } 2$$

$$\Rightarrow \operatorname{corr}(X - \hat{X},\ Y - \hat{Y}) = \text{partial correlation between}$$

X and Y given Z.

PACF (partial autocorrelation function)

- Let p_1 be the correlation between y_t and itself, thus $p_1 = 1$
- p_2 be the partial autocorrelation between y_t and y_{t-2} given y_{t-1} (removing the effect of $y_t - 1$)
- p_3 be the partial autocorrelation between y_t and y_{t-3} given y_{t-1}, y_{t-2} (removing the effect of y_{t-1}, y_{t-2})
- And so on

PACF(k) = p_k

p_2 = partial auto correlation of (y_t, y_{t-2}) given y_{t-1}

PACF of AR(p)

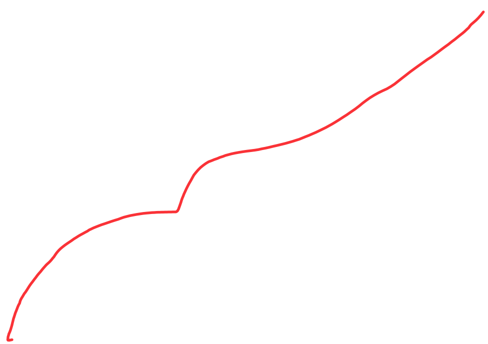
Consider an AR(2) model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon$$

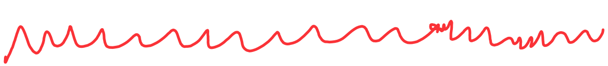
- Then

$$PACF(3) = PACF(4) = \ldots = 0$$

- A time series with non-zeros PACF(2) and zeroes PACF(3), PACF(4)... could be an AR(2) series
- A time series with non-zeros PACF(k) and zeroes PACF(k+1), PACF(k+2)... could be an AR(k) series

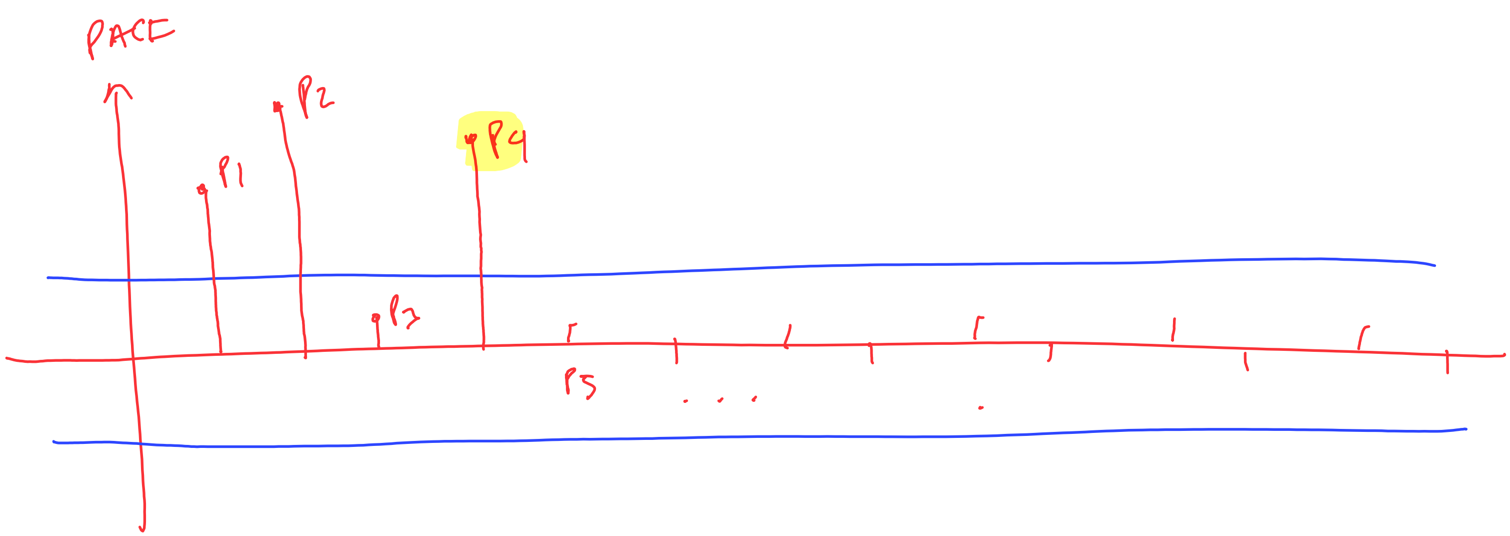


← series with trend ⇒ AR(p) should not be used to model.

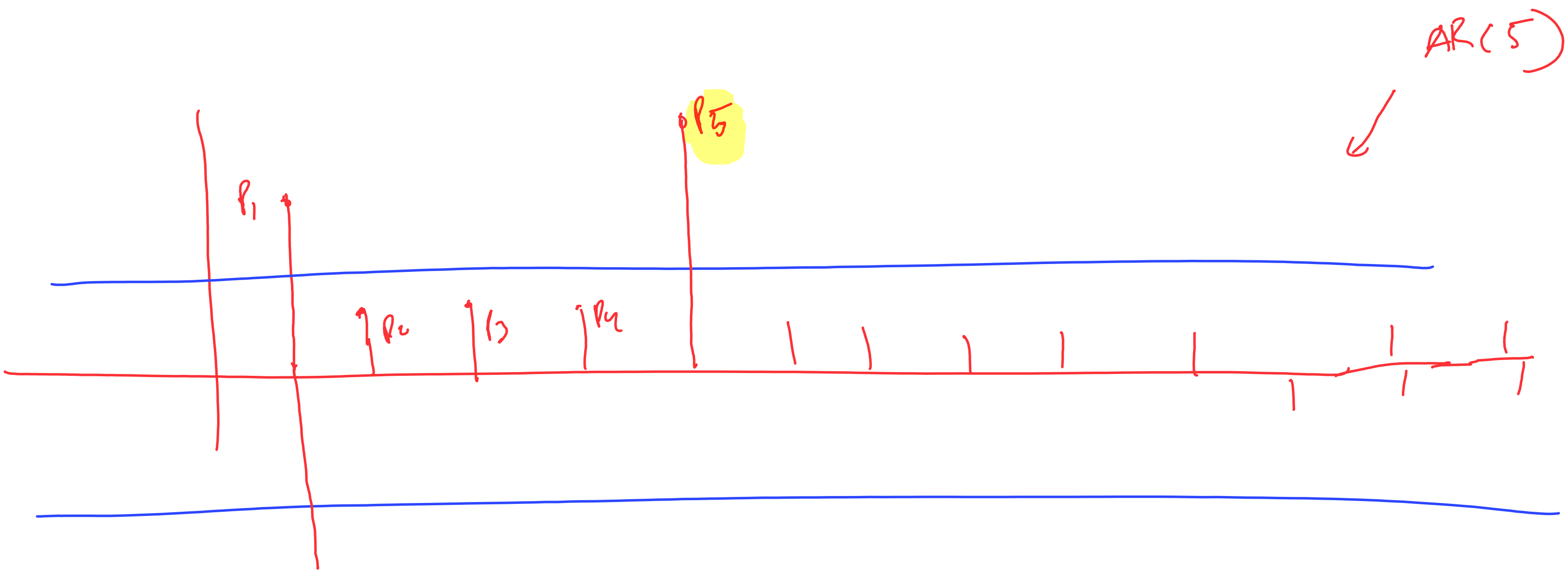


(Stationary series) ⇒ AR(p) can be used to model.

what p value we should select?



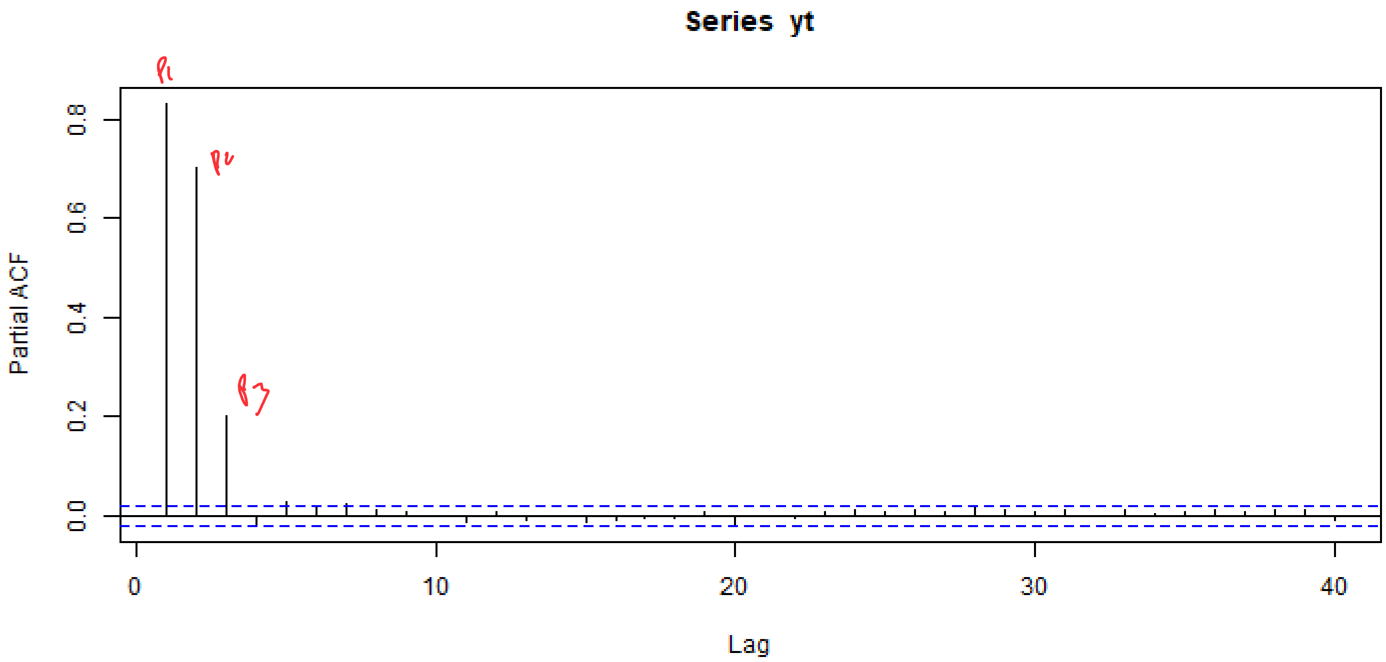
we can use AR(4) to model this series.



Example

```
yt ← arima.sim(list(order=c(3,0,0), ar=c(.1, .65, .2)), n=10000)
b0 = 10
yt ← yt + b0
pacf(yt)
```

AR(3)

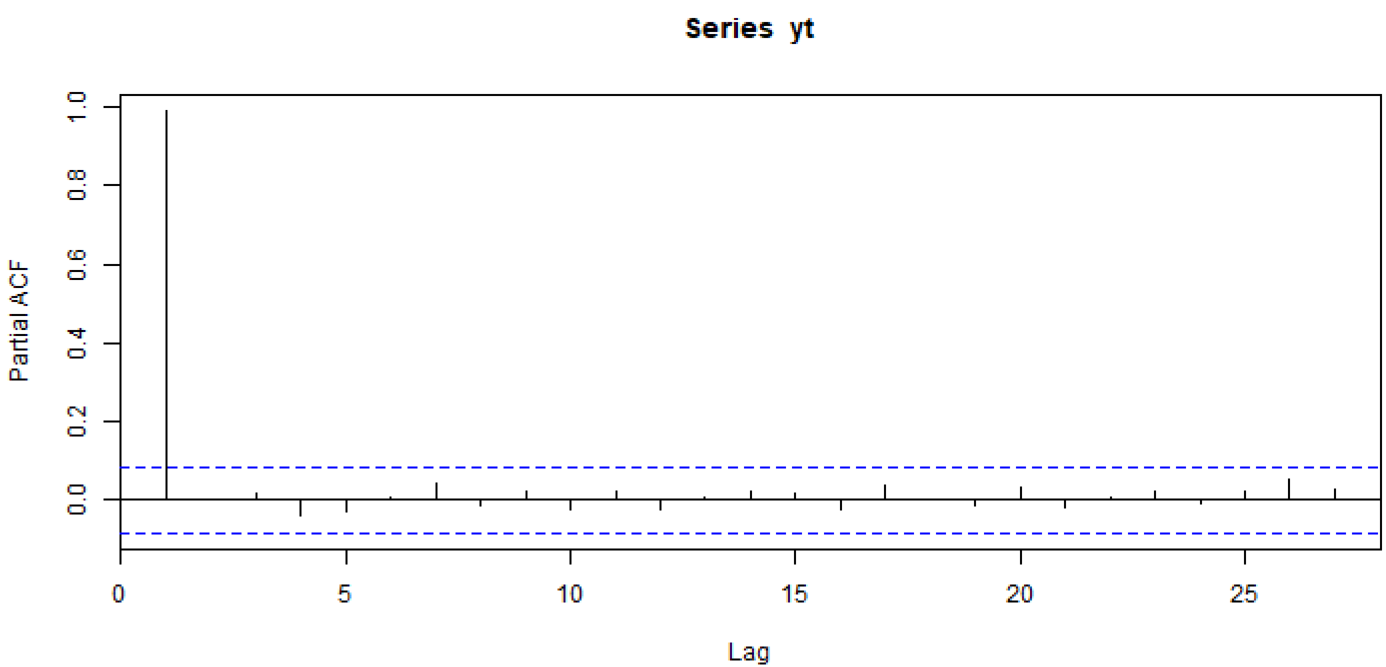


Example

```r
library(quantmod)
library(forecast)
getSymbols("MSFT")
```

```
## [1] "MSFT"
```

```r
yt = MSFT$MSFT.Open
yt <- yt[index(yt) > "2023-01-01"]
pacf(yt)
```



Example

- We notice that PACF(1) is non zeroes and PACF(2), PACF(3)... are zeroes (lie within the blue strip)
- Thus we can use AR(1) to model this series.

Example

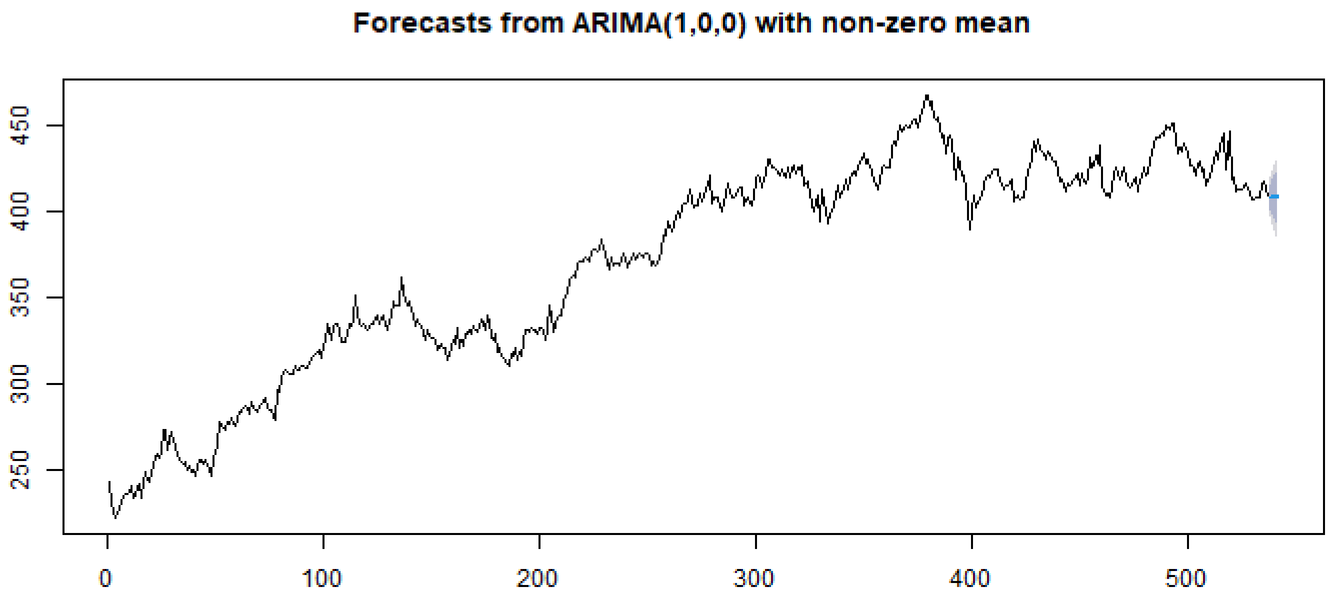
```r
# estimate the series using AR(1) model
yt_ar  = arima(yt, order = c(1,0,0))

# plot the estimated series and the original series
yt_predicted <- yt - yt_ar$residuals

plot(yt)
points(yt_predicted, type = "l",
       col = "red", lty = 2)
```

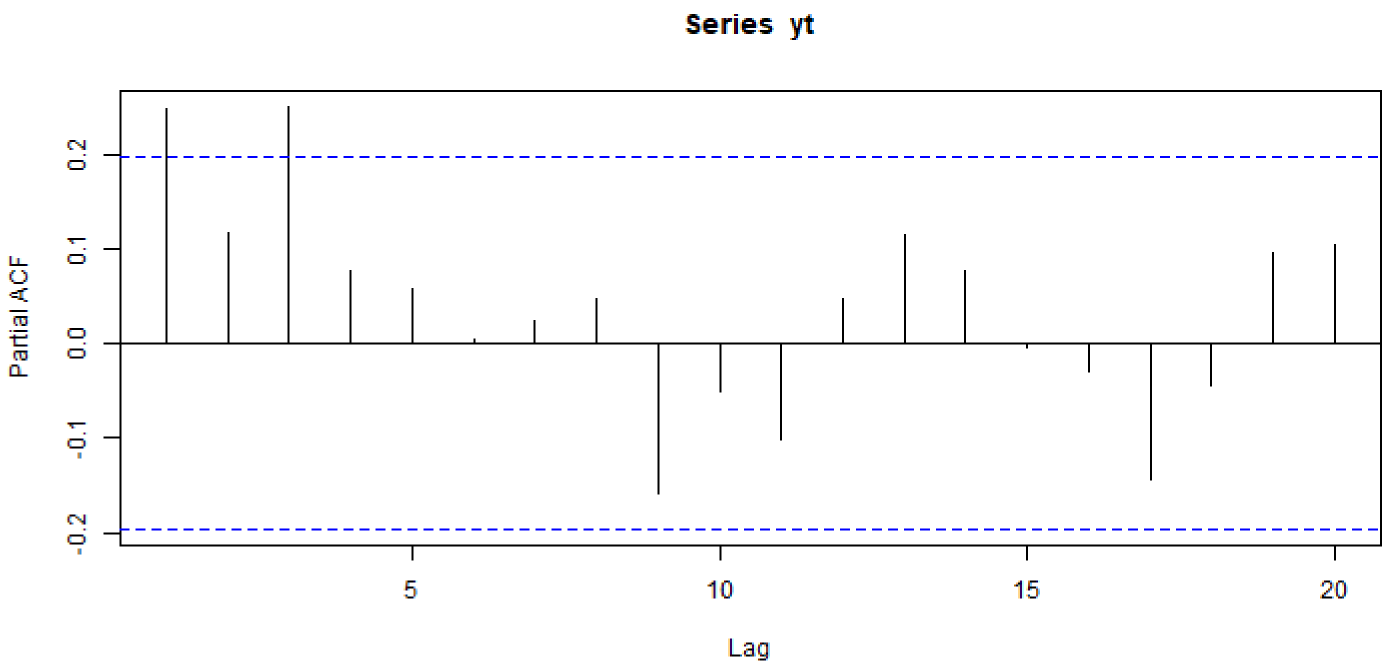
Example

```
yt_forecasts <- forecast(yt_ar, h=5)
plot(yt_forecasts)
```



Example

```r
d = read.csv('earthquakes.csv')
yt = ts(d$Quakes, frequency = 1, start = 1916)
pacf(yt)
```



Example

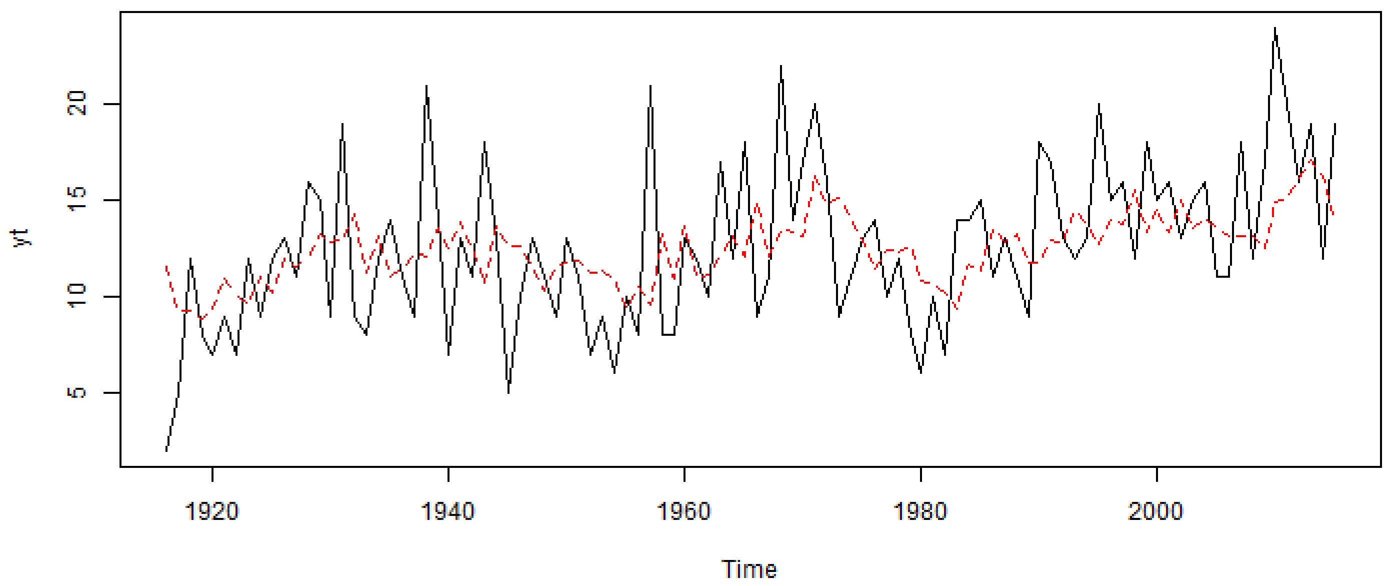
- We notice that PACF(3) is non zeroes and PACF4), PACF(5)... are zeroes (lie within the blue strip)
- Thus we can use AR(3) to model this series.

Example

```r
# estimate the series using AR(1) model
yt_ar  = arima(yt, order = c(3,0,0))

# plot the estimated series and the original series
yt_predicted <- yt - yt_ar$residuals

plot(yt)
points(yt_predicted, type = "l",
       col = "red", lty = 2)
```



Example

```
yt_forecasts <- forecast(yt_ar, h=5)
plot(yt_forecasts)
```