BUAN6357_Homework2_Hou

Chun-Li Hou

2020-09-19

0

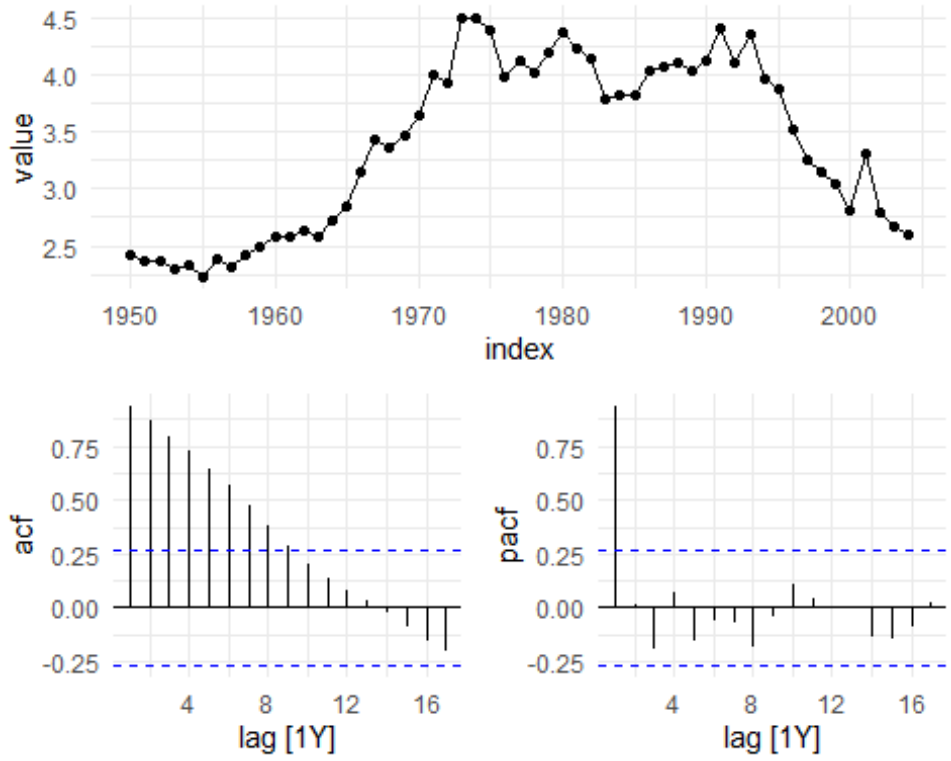
This data set provides information on the number of women murdered each year (per 100,000 standard population) in the U.S.

```
# load package & set environment
if(!require("pacman")) install.packages("pacman")
pacman::p_load(fpp2, fpp3, dplyr, patchwork, purrr)
data("wmurders")
options(digits = 3)
set.seed(42)
theme_set(theme_minimal())
```

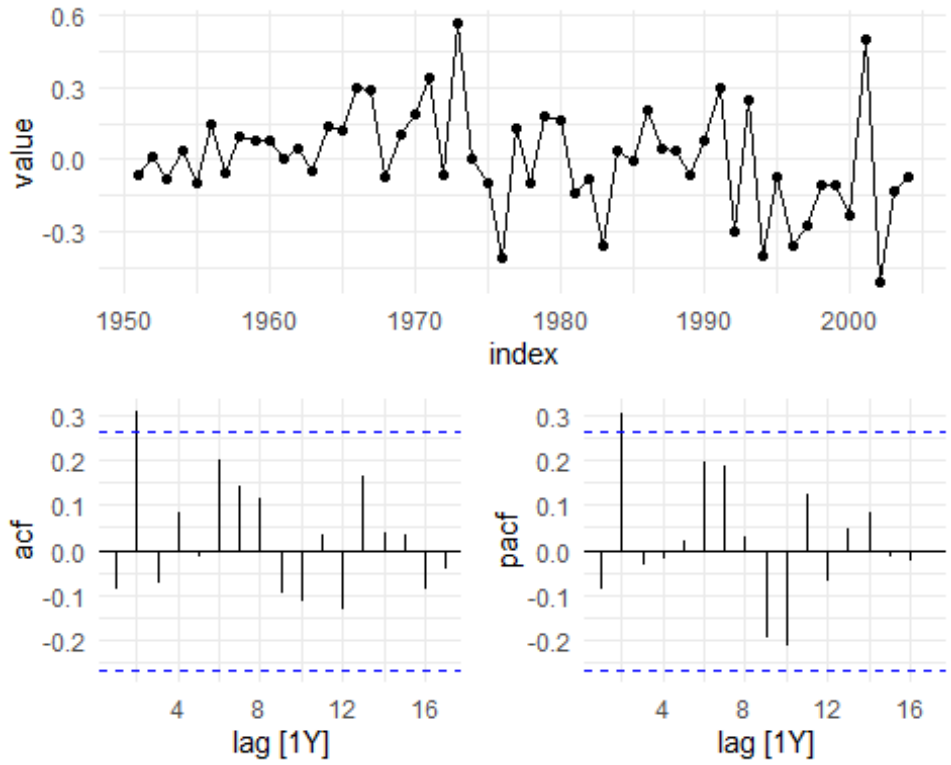
1

By studying appropriate graphs of the series in R, find an appropriate ARIMA(p,d,q) model using first difference (d = 1). If there are two equally likely candidate models, then choose the one with a moving average process (MA).

```
# plot chart
wmurders %>% as_tsibble() %>% gg_tsdisplay(plot_type = "partial")
```



```r
# plot chart
wmurders %>% diff() %>% as_tsibble() %>% gg_tsdisplay(plot_type = "partial")
```

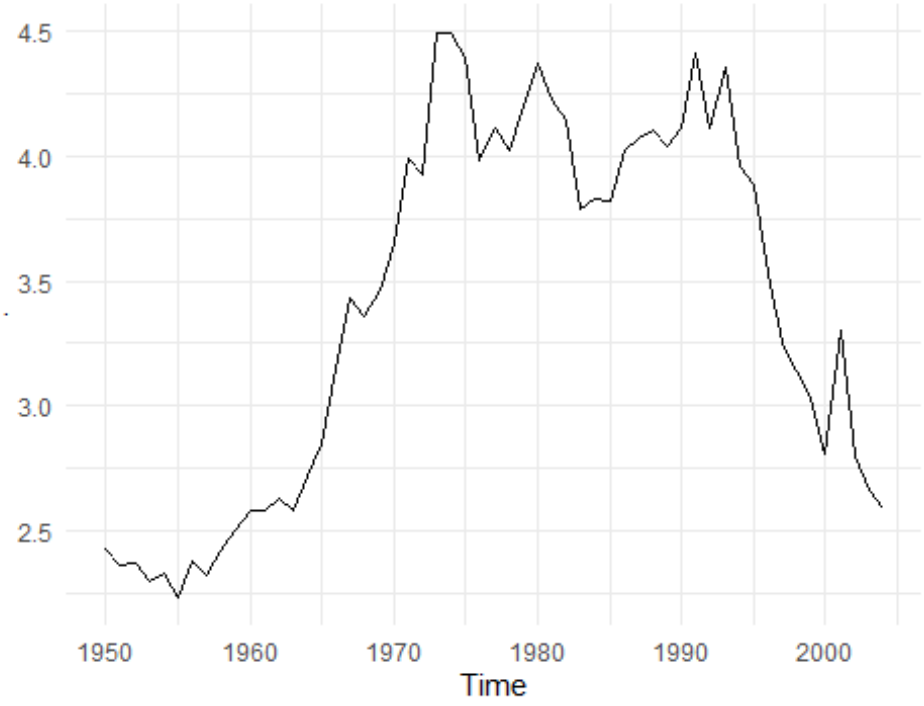


The data at first is clearly not stationary due to having trends. We do the first difference (d = 1) based on the instruction. The result looks better. Also, we decide by the ACF and PACF plot to know that the model could be either p or q as 2. Thus, we have two options But, we choose the one with a moving average process (MA) (q = 2) based on the instruction.

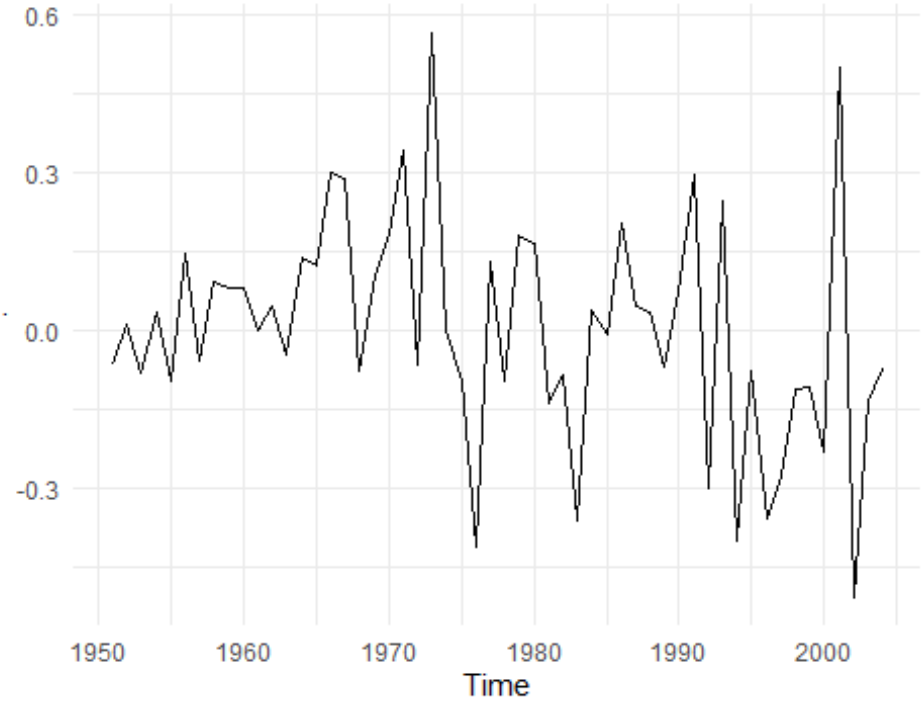
2

Should you include a constant term in the model? Explain your answer.

```
# plot data
wmurders %>% autoplot() # no drift
```



```
wmurders %>% diff() %>% autoplot() # no drift, no trend
```



No, I should not include a constant term in the model because I do not see any significant drift in the plots.

3

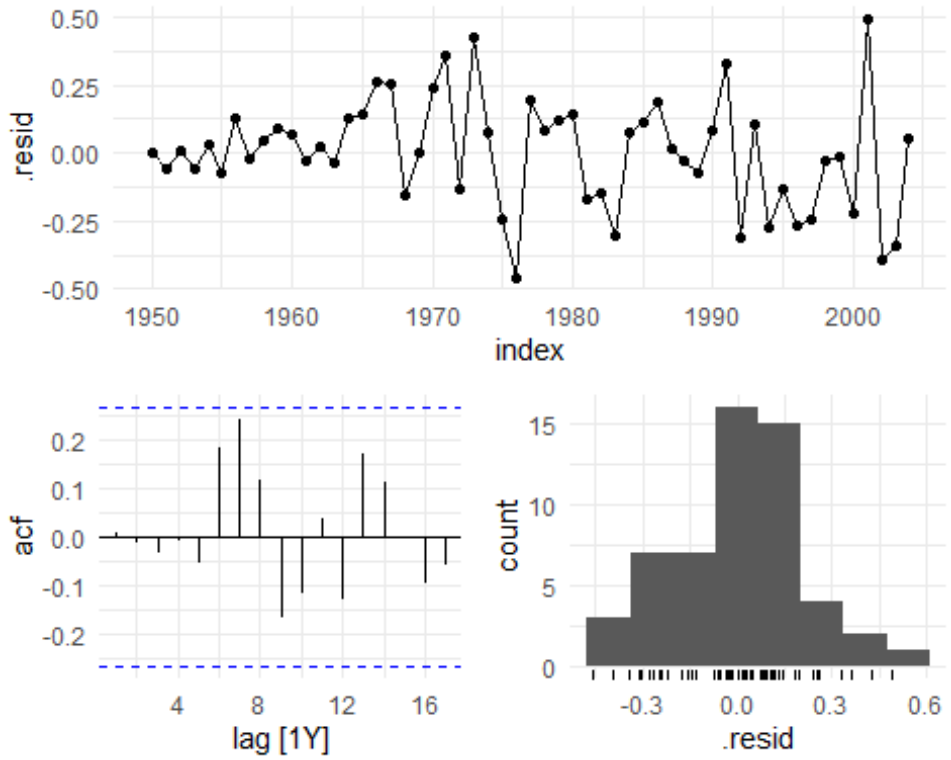
Fit the model using R and examine the residuals. Is the model satisfactory?

```
# fit model
fit = wmurders %>% as_tsibble() %>% model(arima = ARIMA(value ~ pdq(0, 1,
2)))

# print report
fit %>% report()

## Series: value
## Model: ARIMA(0,1,2)
##
## Coefficients:
##          ma1    ma2
##       -0.066  0.371
## s.e.   0.126  0.164
##
## sigma^2 estimated as 0.0422:  log likelihood=9.71
## AIC=-13.4   AICc=-12.9   BIC=-7.46

# check residuals
fit %>% gg_tsresiduals()
```



Yes, the model is satisfactory due to no significant autocorrelation in residuals and distribution of residuals being normal distributed.

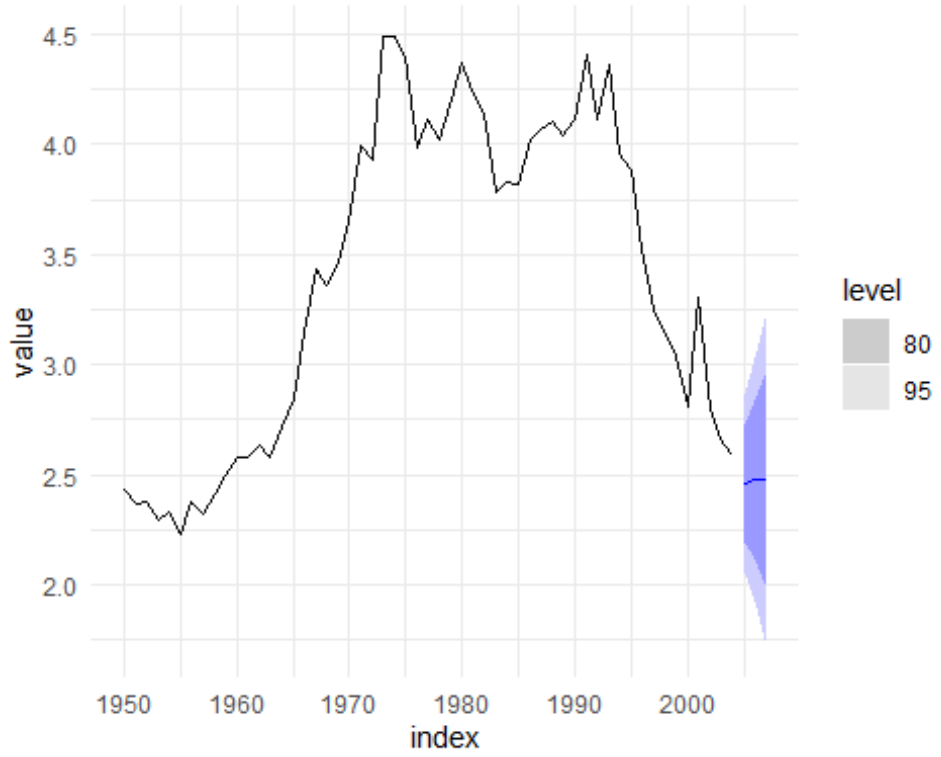
4

Forecast three times ahead and include the results in a table. Also, create a plot of the series with forecasts and prediction intervals for the next three periods shown.

```
# generate forecast
fit %>% forecast(h = 3) %>% print.data.frame()

##   .model index           value .mean
## 1  arima  2005 N(2.46, 0.0422)  2.46
## 2  arima  2006  N(2.48, 0.079)  2.48
## 3  arima  2007  N(2.48, 0.151)  2.48

fit %>% forecast(h = 3) %>% autoplot(wmurders)
```



The table with data and the plot show above.

5

Does ARIMA() give the same model you have chosen? If not, which model do you think is better?

```
# weak auto fit model
fit.auto.w = wmurders %>% as_tsibble() %>%model(ARIMA(value ~ pdq(d = 1)))
fit.auto.w %>% report()

## Series: value
## Model: ARIMA(0,1,0)
##
## sigma^2 estimated as 0.04563:  log likelihood=6.73
## AIC=-11.5   AICc=-11.4   BIC=-9.47

# strong auto fit model
fit.auto.s = wmurders %>% as_tsibble() %>%model(ARIMA(value ~ pdq(d = 1),
                                                      stepwise = F,
                                                      approximation = F))
fit.auto.s %>% report()

## Series: value
## Model: ARIMA(0,1,2)
##
## Coefficients:
##          ma1    ma2
##       -0.066  0.371
## s.e.   0.126  0.164
##
## sigma^2 estimated as 0.0422:  log likelihood=9.71
## AIC=-13.4   AICc=-12.9   BIC=-7.46

# compare result
# fit         AICc=-12.9  ARIMA(0,1,2)
# fit.auto.w  AICc=-11.4  ARIMA(0,1,0)
# fit.auto.s  AICc=-12.9  ARIMA(0,1,2)
```

Yes, the ARIMA() function, in the way that forces to run all combinations (fit.auto.s), gives the same model as chosen before (fit). The one (fit.auto.w) not forced to run all combinations is worse than the one chosen before (fit).