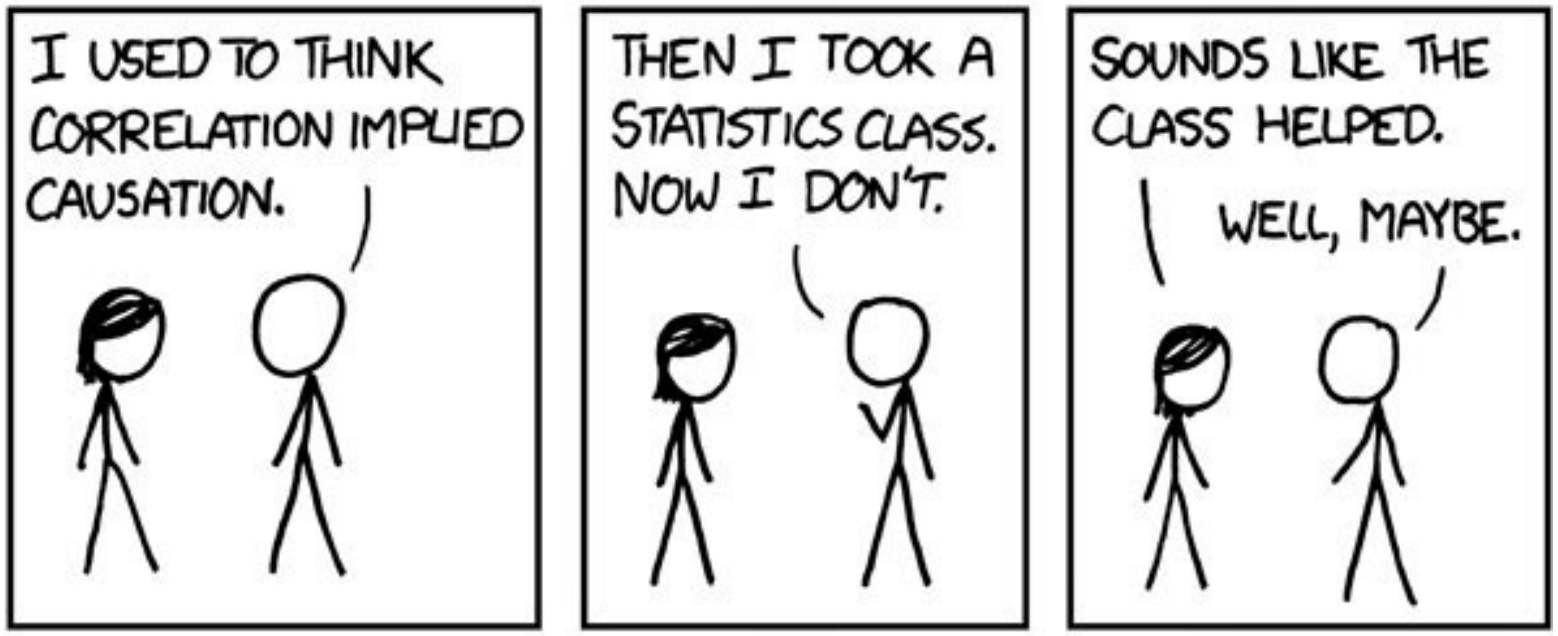

I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.
XKCD

Source: XKCD

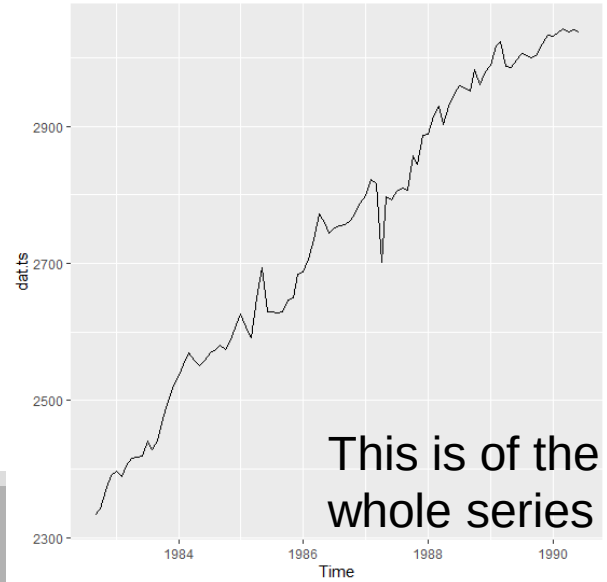
Example

Take a look at the whole process one more time….

Data: consume1982.csv provides monthly disposable spending in 1982 dollars (the units are in billions of dollars!!)

The data starts in September 1982 and goes to June 1990

- Training data set consists of 70 observations (September 1982 – June 1988)
- Validation data set consists of 12 observations (July 1988 – June 1989)
- Test data set consists of 12 observations (July 1989 – June 1990)



Looks like it is a Random Walk with drift

```
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -4.40  0.0100
[2,]   1 -3.66  0.0347
[3,]   2 -3.02  0.1585
```

Creating Models

Going to create a model using auto.arima (Model 1) then one on my own (Model2)

Using auto.arima….

ARIMA(0,1,1)

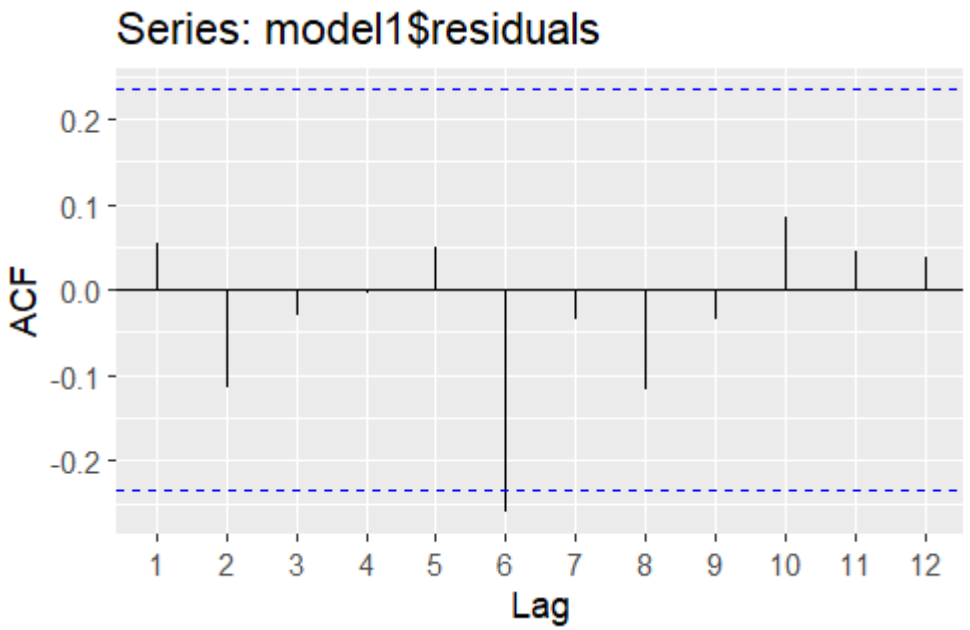
Coefficients:

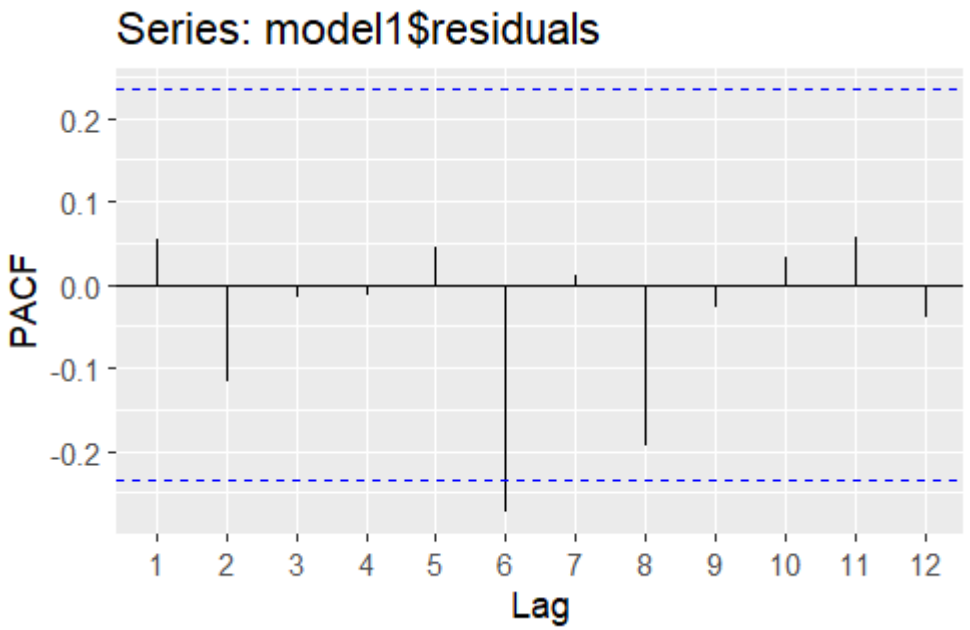
	ma1	drift
	-0.4362	8.7689
s.e.	0.1298	1.6818

sigma^2 estimated as 617.5: log likelihood=-318.68
AIC=643.36   AICc=643.73   BIC=650.06

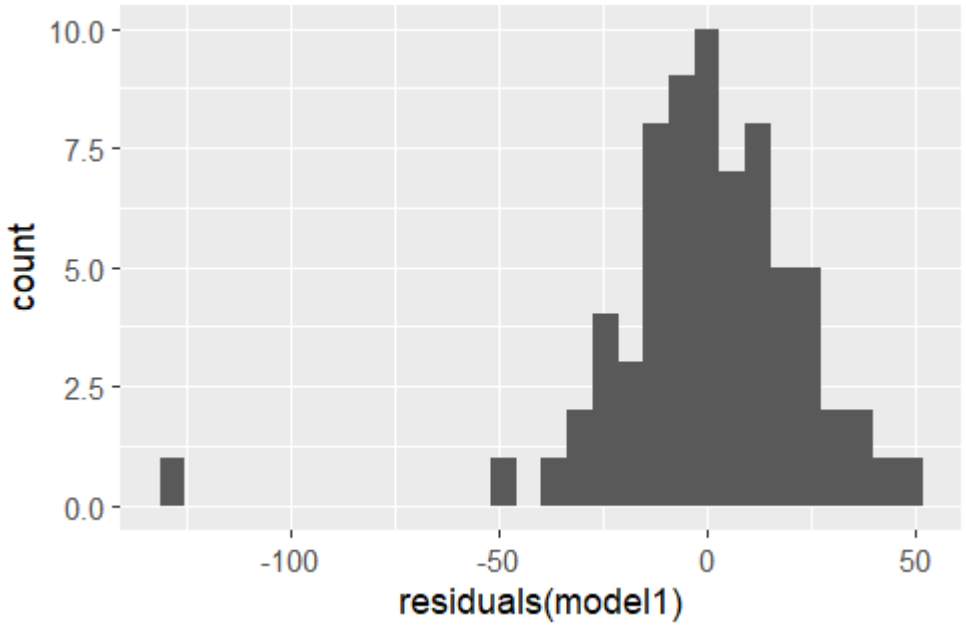
MAE	MPE	MAPE
16.604	0.002439733	0.6241532

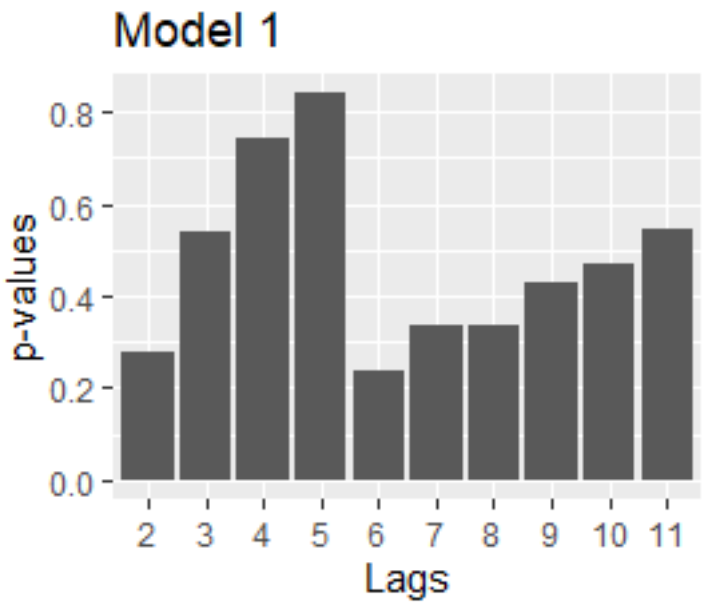
Didn’t do bad, but potentially there are still lags



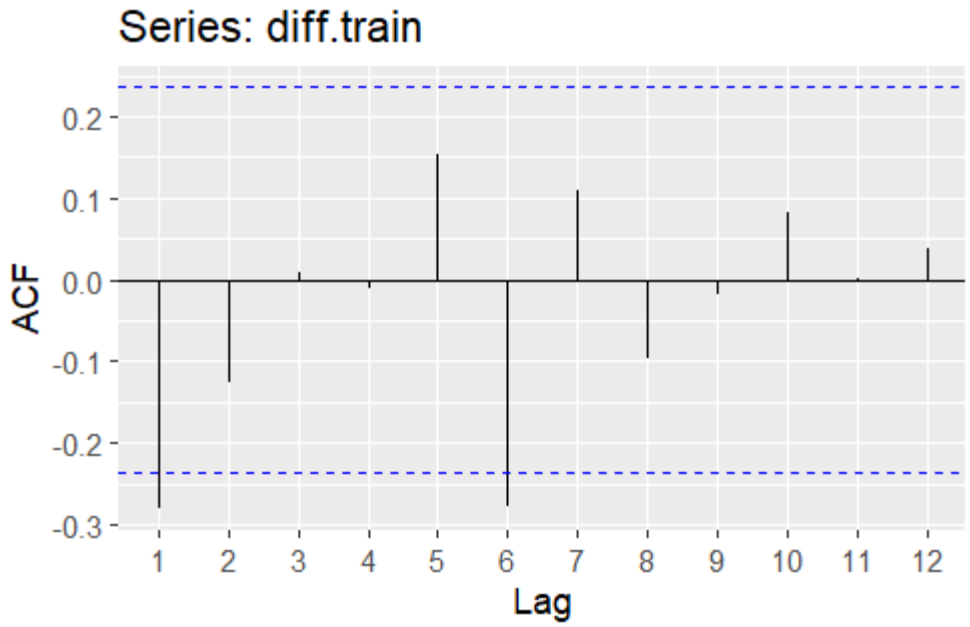


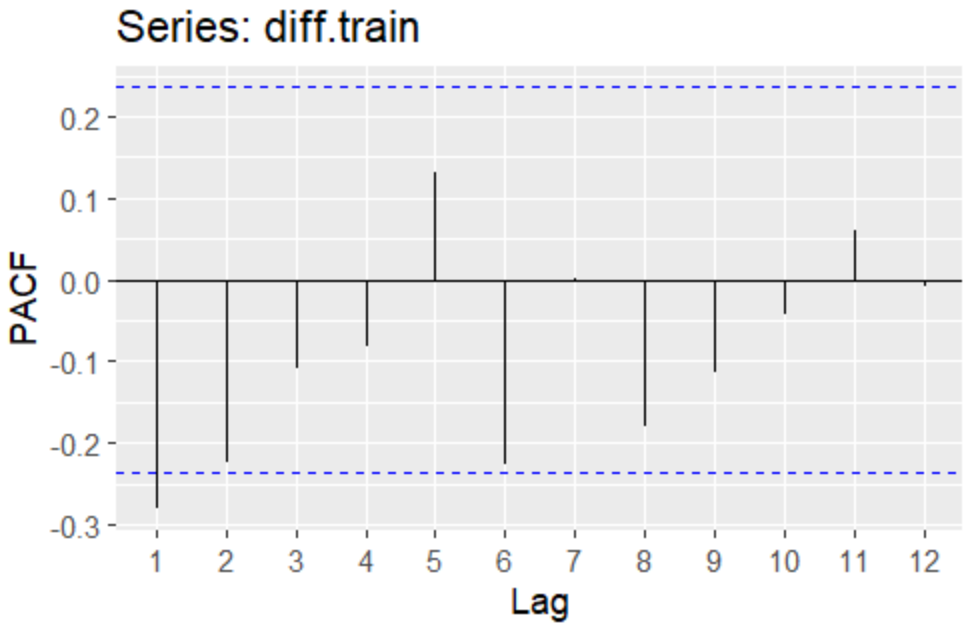
Not bad for white noise (alas, do have a pretty big outlier…the big dip down in April of 1987)





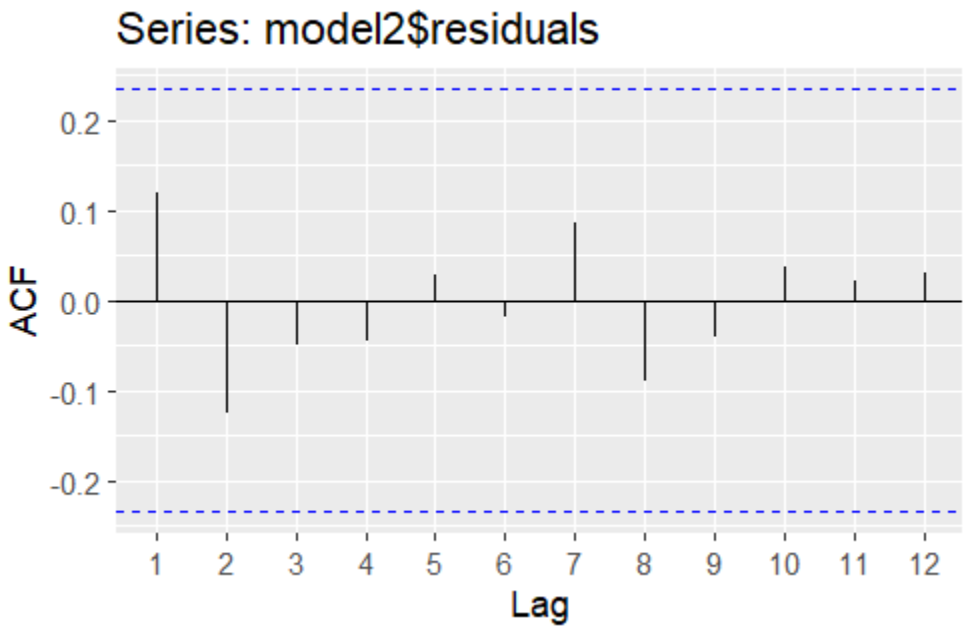
Now try one by hand…first look at ACF and PACF of differences…

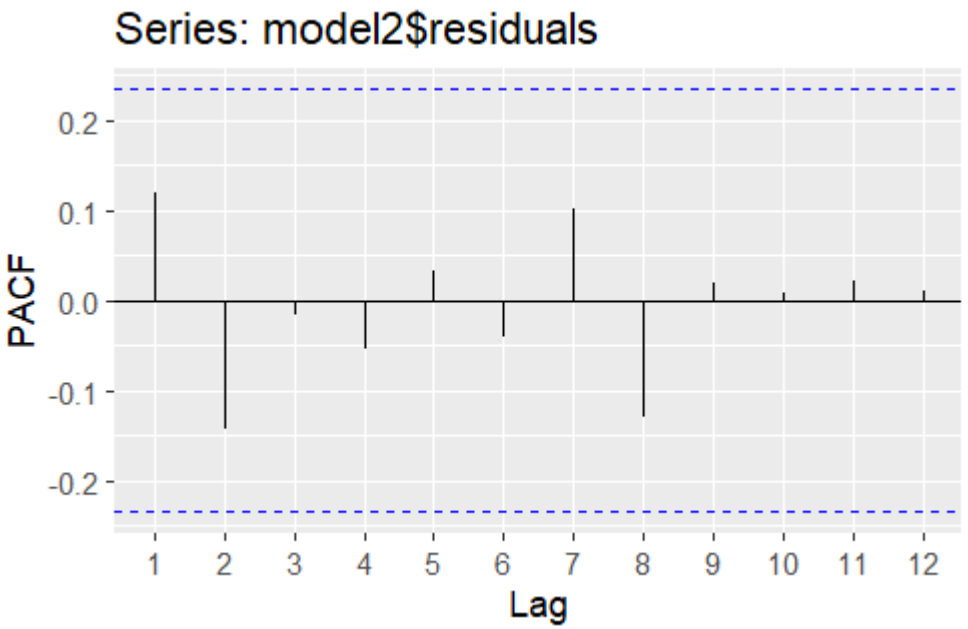


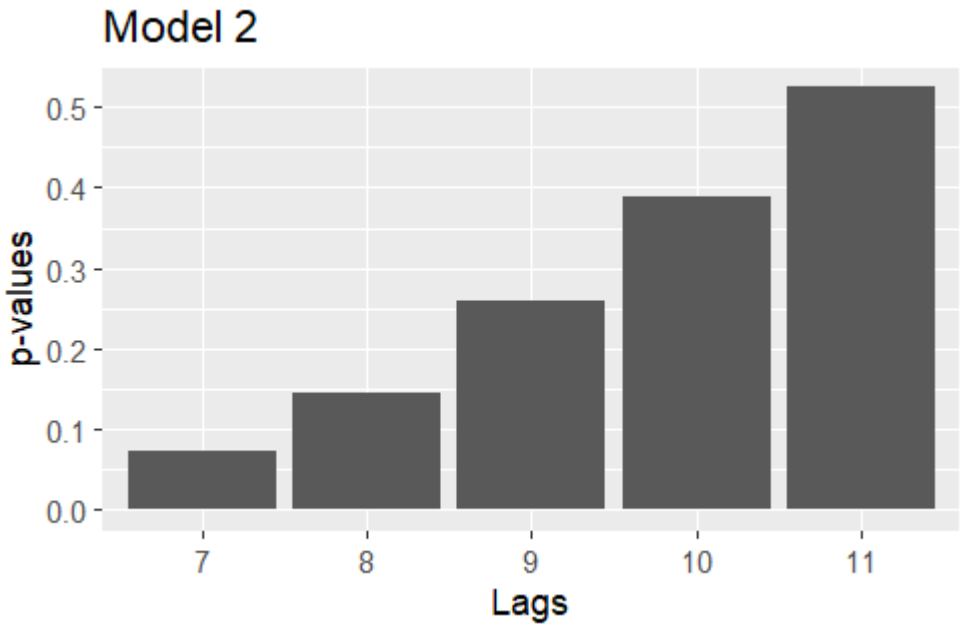


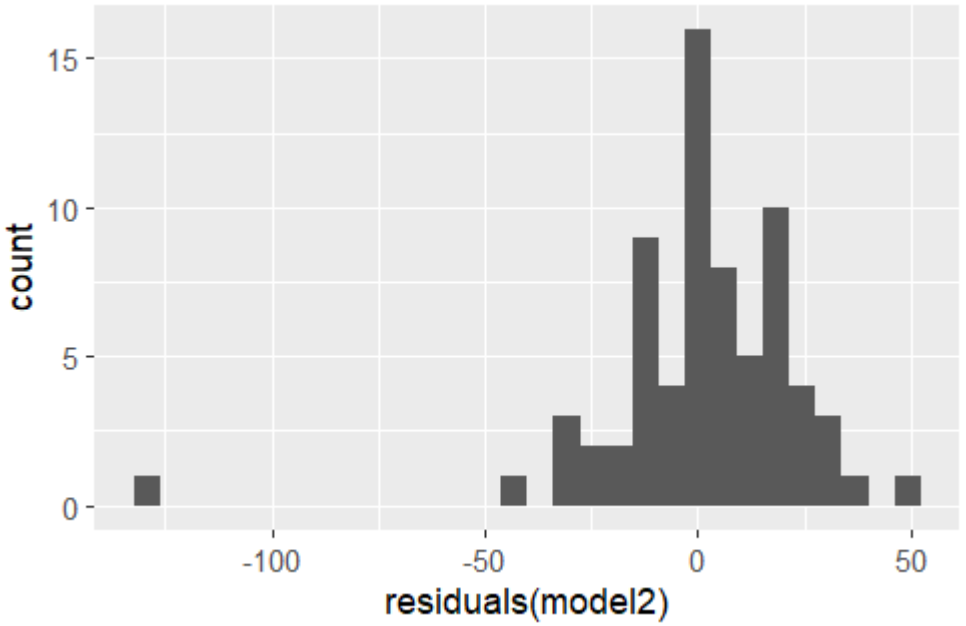
```
model2=Arima(train.ts,order=c(0,1,6),include.drift = T, fixed=c(NA,0,0,0,0,NA,NA))


Series: train.ts
ARIMA(0,1,6) with drift

Coefficients:
          ma1      ma2  ma3  ma4  ma5      ma6      drift
      -0.5556    0      0       0       0    -0.2935    8.2303
s.e.   0.1763    0      0       0       0     0.1083    0.6218

sigma^2 estimated as 562.1:  log likelihood=-315.5
AIC=638.99   AICc=639.62   BIC=647.93
```

MAE	MPE	MAPE
15.08845	0.05253814	0.5687325


Series: model2$residuals
ACF
0.2
0.1
0.0
-0.1
-0.2
1
2
3
4
5
6
7
8
9
10
11
12
Lag


Series: model2$residuals
PACF
0.2
0.1
0.0
-0.1
-0.2
1
2
3
4
5
6
7
8
9
10
11
12
Lag


Model 2
p-values
0.0
0.1
0.2
0.3
0.4
0.5
7
8
9
10
11
Lags


count
0
5
10
15
-100
-50
0
50
residuals(model2)

MAPE.1	MAE
0.006836268	20.3882
MAPE.2	MAE
0.005492793	16.4175