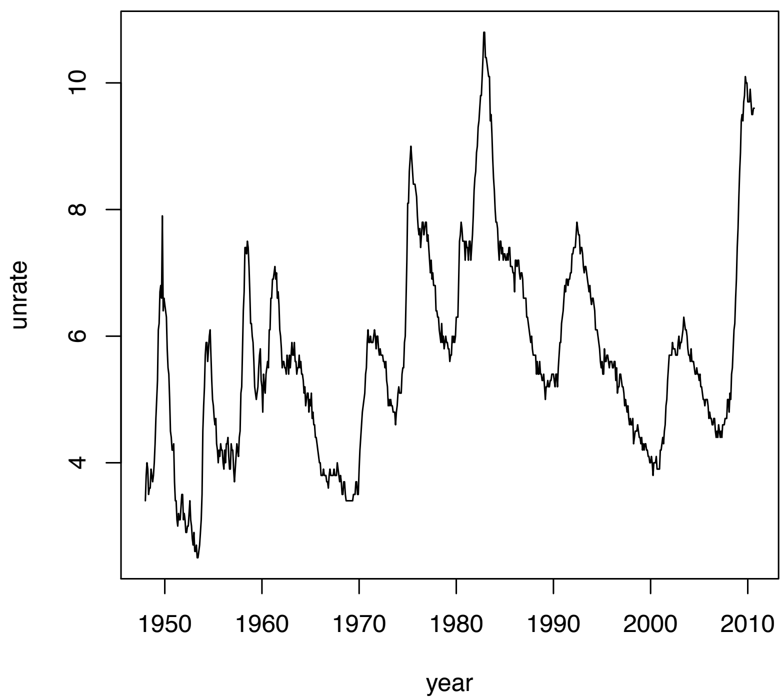
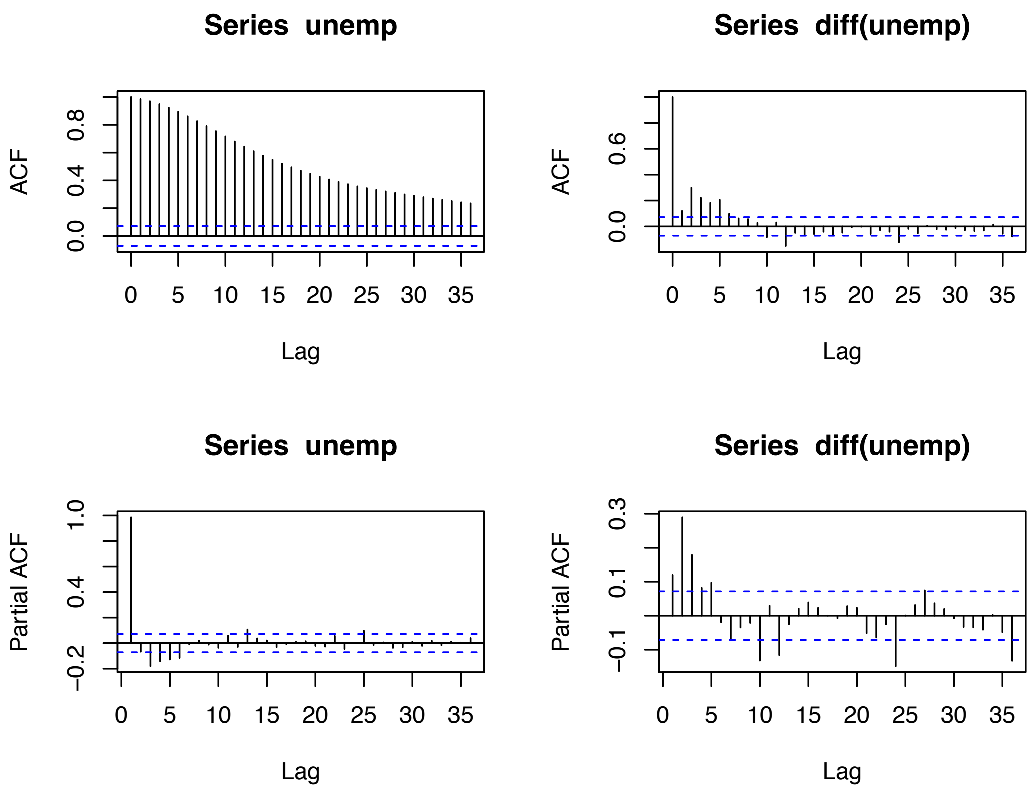
Seasonal ARIMA models and regression with time series errors

US monthly unemployment rates (seasonal ARIMA)

- Data: 1948 to 2010 monthly unemployment rates
- Strong cyclical patterns, upward trend and asymmetry on rise and decline.



- acf and pacf of the original data and differenced data



Model identification

- Sample ACF of the original data is persistent, reflecting the upward trend of the data
- Sample ACF and PACF of the differenced data are small and decays quickly.
- Significant lags at 12, 24 and 36 for both acf and pacf of the differenced data.
  Indicating that seasonality is not completely removed.
- The acf and pacf at seasonal lags do not cut off.
- Both acf and pacf of the differenced data have first 5 lags being significant. PACF is roughly exponential decaying. Thus we entertain the model with p=1 and q=5.
- The model is

$$(1-\Phi B^{12})(1-\phi B)(1-B)X_t = (1-\theta_1 B-\cdots-\theta_5 B^5)(1-\Theta B^{12})w_t.$$

Fitted models

The fitted model is

$$
\begin{aligned}
& (1-0.6B^{12})(1-0.73B)(1-B)X_t \\
= \; & (1-0.75B+0.22B^2+0.01B^3+0.04B^4+0.08B^5)(1-0.85B^{12})w_t.
\end{aligned}
$$

with $\sigma^2 = 0.03643$.

```
Call:
arima(x = unemp, order = c(1, 1, 5),
seasonal = list(order = c(1, 0, 1), period = 12))

Coefficients:
         ar1      ma1     ma2     ma3     ma4     ma5    sar1     sma1
      0.7301  -0.7468  0.2194  0.0073  0.0383  0.0831  0.5978  -0.8469
s.e.  0.0686   0.0776  0.0462  0.0501  0.0467  0.0431  0.0672   0.0477

sigma^2 estimated as 0.03643:  log likelihood = 176.43,  aic = -334.87
```

Improved model

The MA coefficients 3 and 4 are not significant. Thus we run the reduced model and get

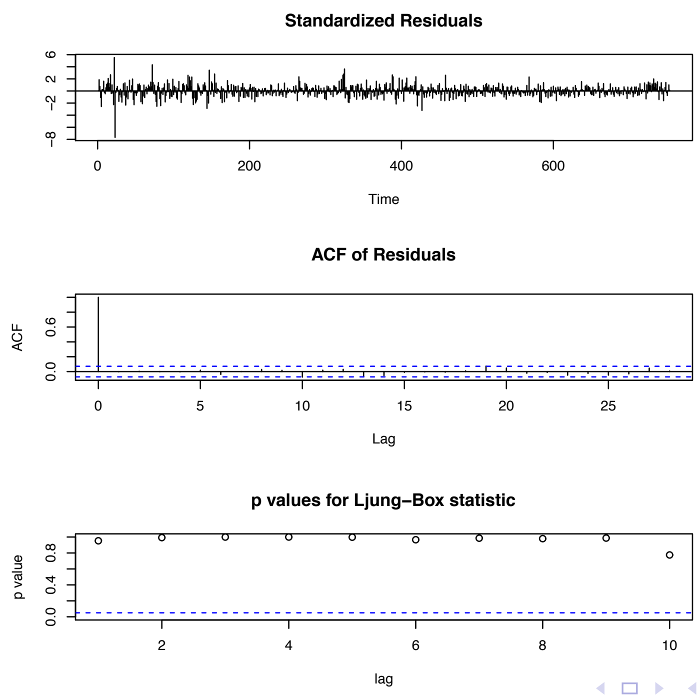
$$\begin{aligned} & (1-0.61B^{12})(1-0.75B)(1-B)X_t \\ = \; & (1-0.77B+0.24B^2+0.099B^5)(1-0.85B^{12})w_t. \end{aligned}$$

with $\sigma^2 = 0.03649$.

```
Call:
arima(x = unemp, order = c(1, 1, 5), seasonal = list(order = c(1, 0, 1), period = 12),
    fixed = c(NA, NA, NA, 0, 0, NA, NA, NA))

Coefficients:
         ar1      ma1     ma2  ma3  ma4     ma5    sar1     sma1
      0.7536  -0.7744  0.2351    0    0  0.0990  0.6051  -0.8525
s.e.  0.0569   0.0650  0.0365    0    0  0.0386  0.0654   0.0457

sigma^2 estimated as 0.03649:  log likelihood = 175.75,  aic = -337.5
```

TSdiag for model 1

- The tsdiag plots show that both models are adequate. Except for a few outliers at the begining the residual plot also seems good.
- The second model has a smaller AIC score.



TSdiag for model 2

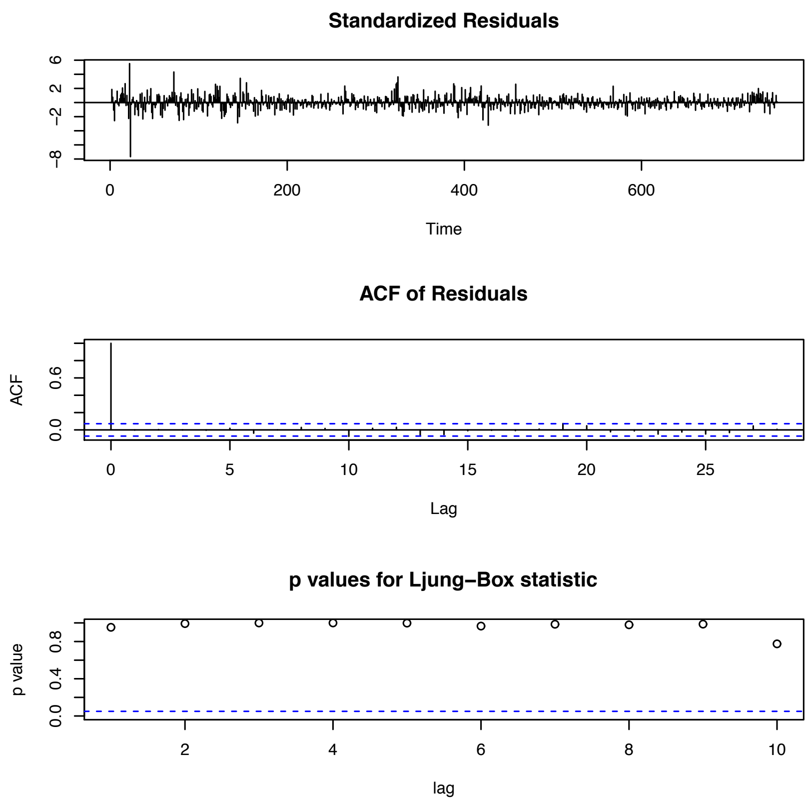


Figure: ACF and PACF

Box tests

Box tests at 36 lags show residuals are uncorrelated.

```
> Box.test(m1$residuals,lag=36,type='Ljung')

        Box-Ljung test

data:  m1$residuals
X-squared = 31.3794, df = 36, p-value = 0.688

> Box.test(m2$residuals,lag=36,type='Ljung')

        Box-Ljung test

data:  m2$residuals
X-squared = 32.4586, df = 36, p-value = 0.6378
```

code

```
da=read.table("m-unrate.txt",header=T)
dim(da)
unemp=da$rate
unrate=ts(unemp,frequency=12,start=c(1948,1))
plot(unrate,xlab='year',ylab='unrate',type='l')

par(mfcol=c(2,2))
acf(unemp,lag=36)
pacf(unemp,lag=36)
acf(diff(unemp),lag=36)
pacf(diff(unemp),lag=36)

m1=arima(unemp,order=c(1,1,5),seasonal
=list(order=c(1,0,1),period=12))

m2=arima(unemp,order=c(1,1,5),seasonal=list(order=
c(1,0,1),period=12),fixed=c(NA,NA,NA,0,0,NA,NA,NA))

tsdiag(m1)
tsdiag(m2)

Box.test(m1$residuals,lag=36,type='Ljung')
Box.test(m2$residuals,lag=36,type='Ljung')
```

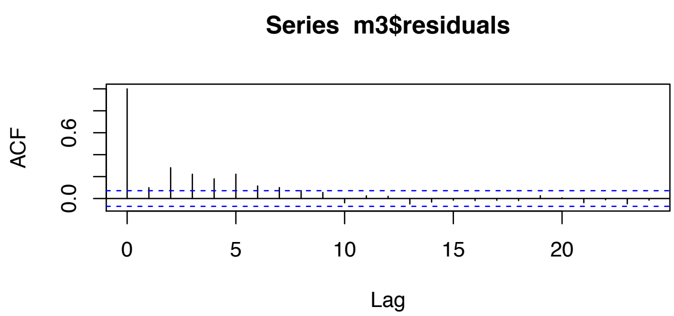
Alternative model specification

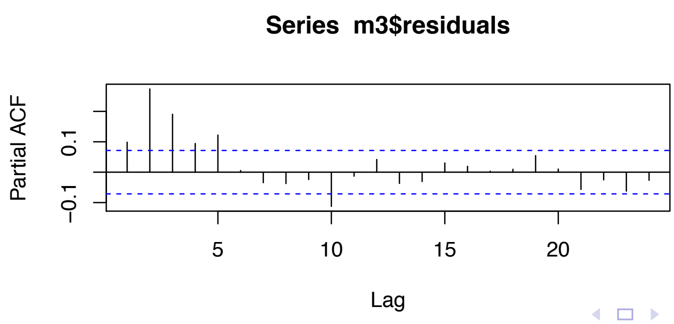
- Two steps procedure for model specification.
- First, focus on the seasonal pattern by looking at the ACF and PACF of the differenced data.
- According to previous analysis, the seasonal pattern can be eliminated via seasonal ARMA(1,1) model.
- We fit $(1 - \Phi B^{12})(1 - B)X_t = (1 - \Theta B^{12})w_t$. The fitted model is

$$(1 - 0.62B^{12})(1 - B)X_t = (1 - 0.87B^{12})w_t$$

Examing residuals

- The seasonal ARMA(1,1) model is not adequate. But judging from the residual plot, the seasonal effects are largely gone.
- A candidate model for the residual is AR(5) as the pacf seems to cut off at lag 5.





Fit a combined model

- Now we fit a combined model $(5, 1, 0) \times (1, 0, 1)_{12}$.
- The first lag is not significant.
- Fixed this term to zero.

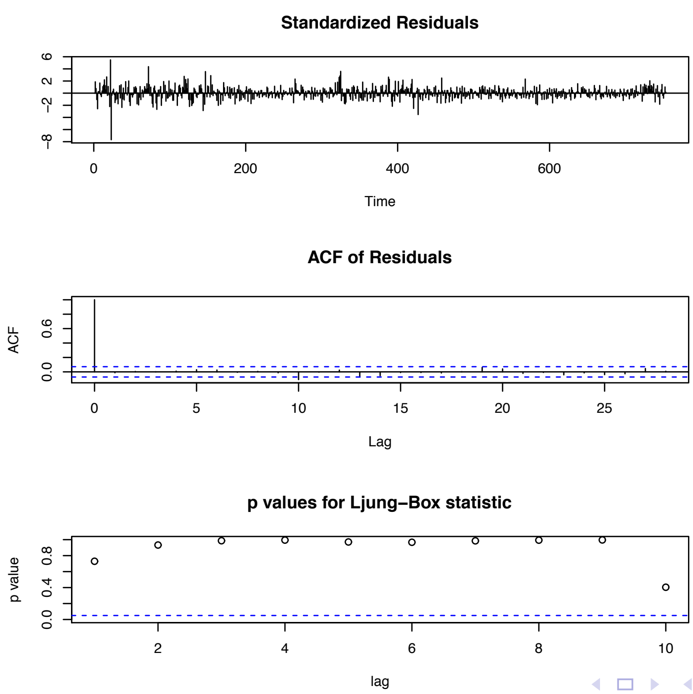
```
Coefficients:
          ar1     ar2     ar3     ar4     ar5    sar1     sma1
      -0.0124  0.2101  0.1682  0.1024  0.1207  0.5624  -0.8233
s.e.   0.0365  0.0366  0.0366  0.0370  0.0366  0.0723   0.0526

sigma^2 estimated as 0.03663:  log likelihood = 174.57,
  aic = -333.13
```

Fitted model

The fitted model is

$$\begin{aligned} & (1 - 0.21B^2 - 0.17B^3 - 0.1B^4 - 0.12B^5)(1 - 0.56B^{12})(1 - B)X_t \\ = \quad & (1 - 0.82B^{12})w_t. \end{aligned}$$

with $\sigma^2 = 0.037$.



Back test

- The fitted model 5 seems to be adequate.
- We perform backtest using model 2 and model 5.

```
> test1=backtest(m1,unemp,700,1,fixed=c1,inc.mean=F)
[1] "RMSE of out-of-sample forecasts"
[1] 0.1662391
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.1349363
> test1=backtest(m4,unemp,700,1,fixed=c2,inc.mean=F)
[1] "RMSE of out-of-sample forecasts"
[1] 0.1679285
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.1350412
```

code

```
m3=arima(unemp, order=c(0,1,0), seasonal=list(order=c(1,0,1), period =12))
m3

par(mfcol=c(2,1))
acf(m3$residuals,lag=24)
pacf(m3$residuals, lag=24)

m4=arima(unemp, order=c(5,1,0), seasonal=list(order=c(1,0,1), period =12))
m4

c1=c(NA,NA,NA,0,0,NA,NA,NA)
c2=c(0,NA,NA,NA,NA,NA,NA)

m5=arima(unemp, order=c(5,1,0), seasonal=list(order=c(1,0,1), period =12),fixed=c2)
tsdiag(m5)

test1=backtest(m1,unemp,700,1,fixed=c1,inc.mean=F)
test1=backtest(m4,unemp,700,1,fixed=c2,inc.mean=F)
```