

Bruce Campell NCSU ST 534 HW 4

Problems 3.32, 3.35(a), and 3.43

Shumway, Robert H.; Stoffer, David S. Time Series Analysis and Its Applications: With R Examples (Springer Texts in Statistics)

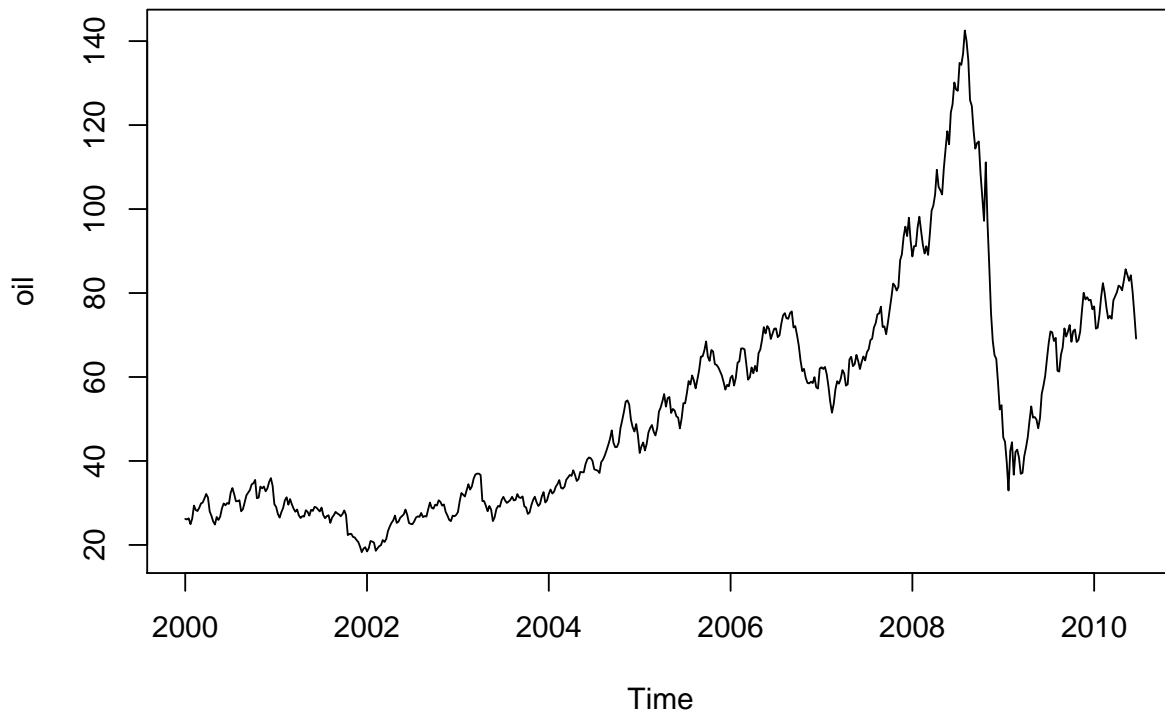
10 November, 2017

3.32 oil time series analysis

Crude oil prices in dollars per barrel are in oil; see Appendix R for more details. Fit an $ARIMA(p, d, q)$ model to the growth rate performing all necessary diagnostics. Comment.

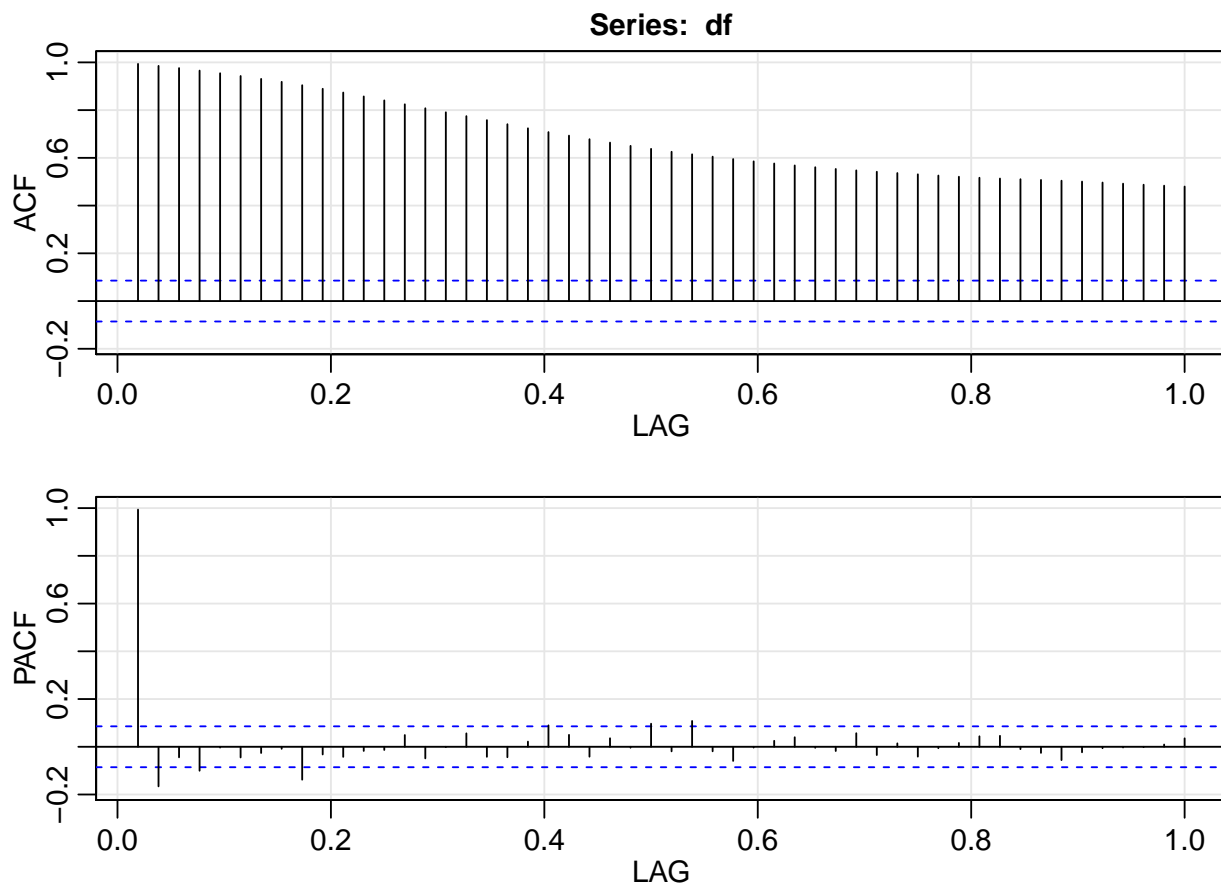
```
library(astsa)
data(oil, package = "astsa")
df <- oil
plot(oil, main = "Crude oil, WTI spot price FOB (in dollars per barrel) weekly")
```

Crude oil, WTI spot price FOB (in dollars per barrel) weekly



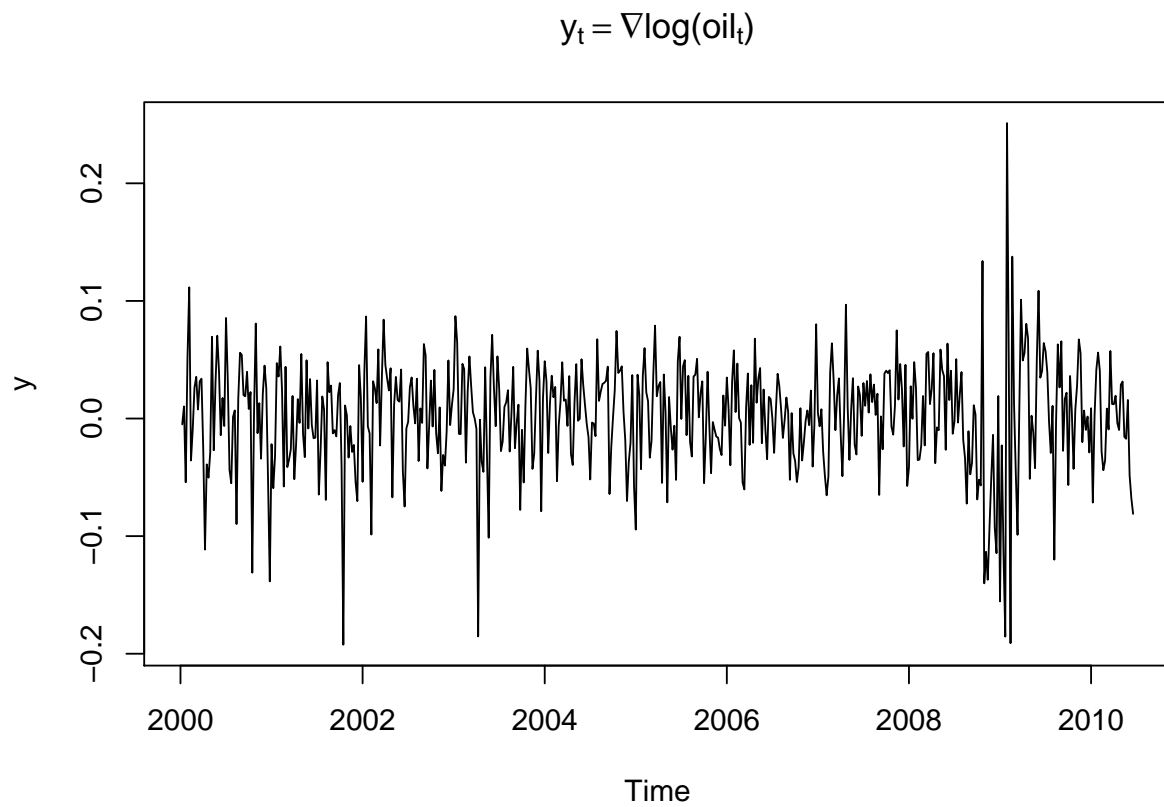
Here is the acf of the oil series.

```
acf2(df, 52)
```



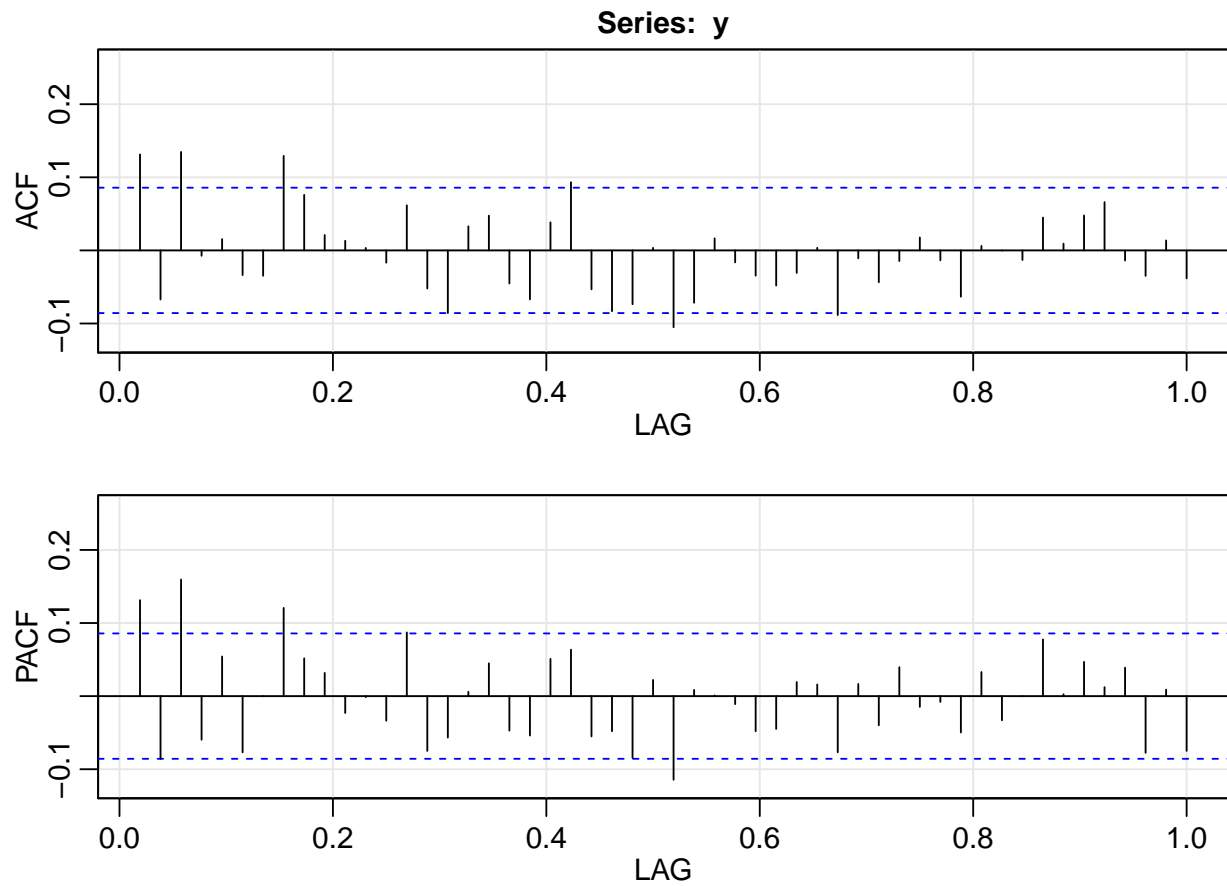
As expected the trend obscures the underlying structure of the fluctuations. We will now calculate $y_t = \nabla \log(x_t)$ and display the ACF and PACF.

```
y <- diff(log(df))  
plot(y, main = TeX("$y_t = \\nabla \log(oil_t)$"))
```



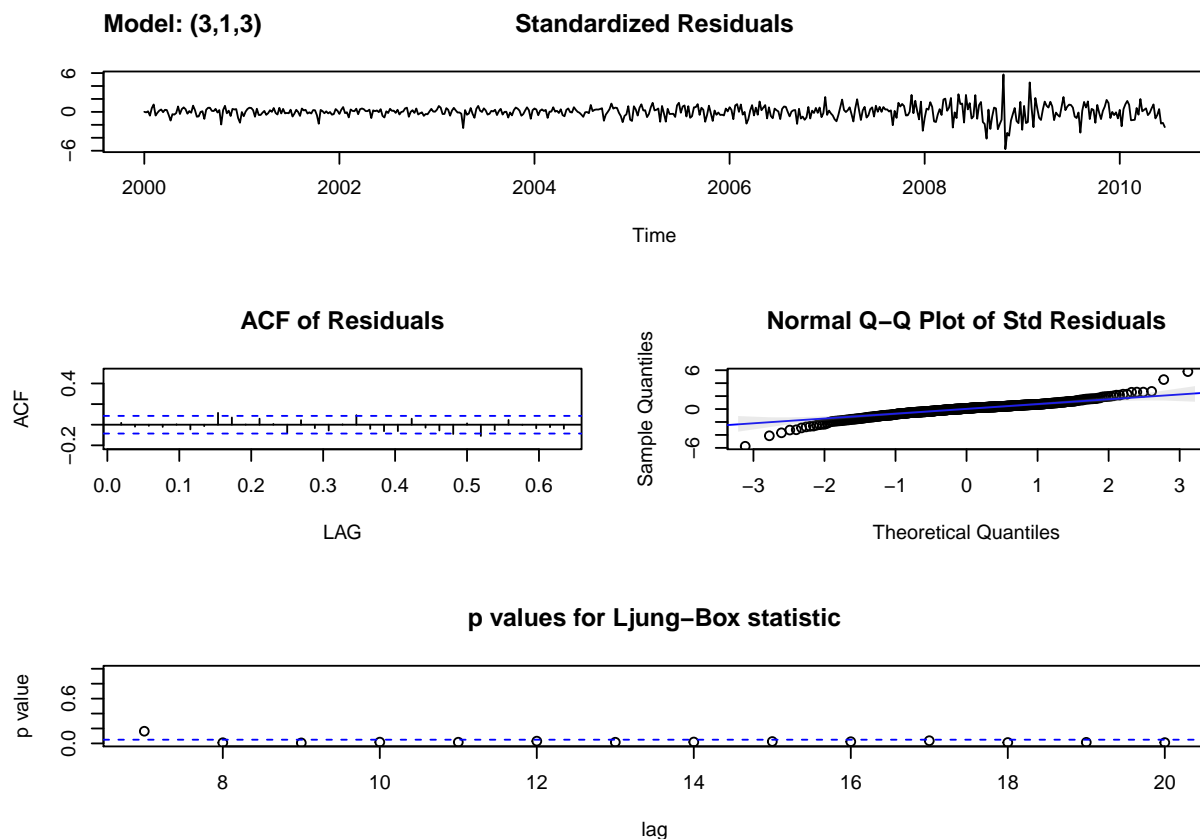
We now see the underlying structure better. There's a period around 2008-2009 that one could argue requires more sophisticated modelling such as stochastic volatility.

```
acf2(y, 52)
```



Based on the ACF and PACF of the differenced log series - we will try to fit an $ARIMA(3, 1, 3)$ model to $\log(x_t)$

```
invisible(model <- sarima(df, 3, 1, 3))
```



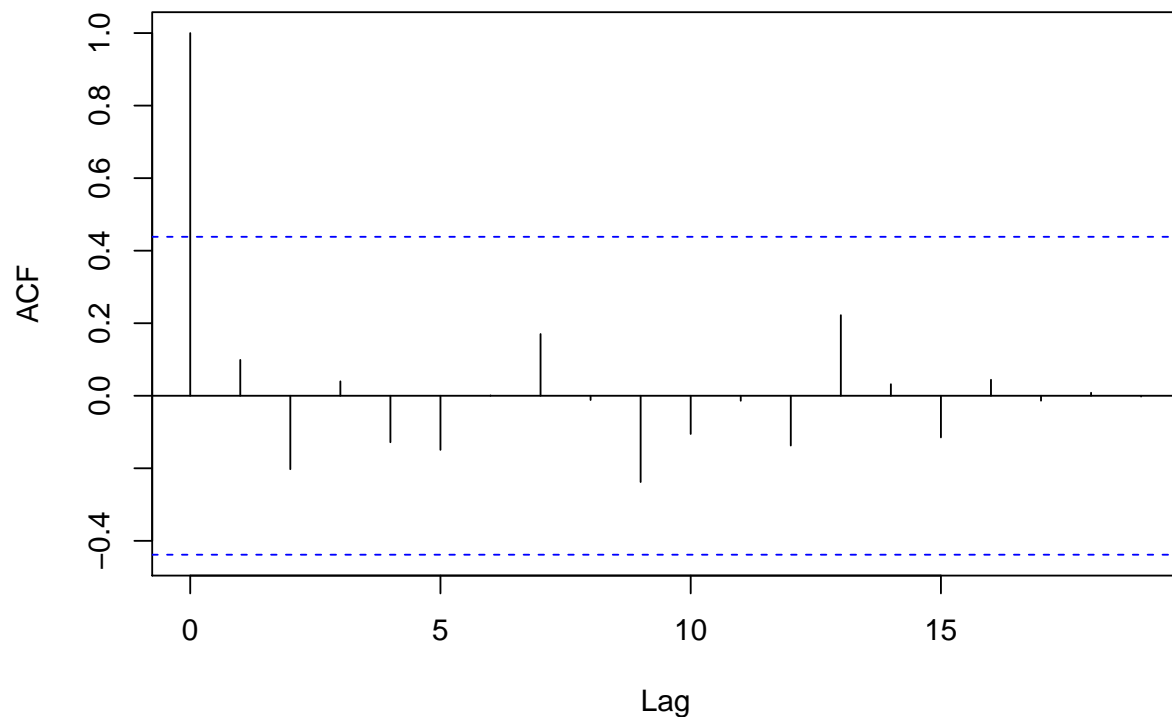
```
model$tttable
```

For lag 1 the Ljung-Box statistic shows significant correlation in the residuals.

For fun let's calculate the Ljung-Box-Pierce Q-statistic to check for systemic autocorrelation in the residuals. We'll extract the residuals and do the calculation by hand - there's an R function for this `Box-Test` that we've experimented with. we'll revisit this.

```
n <- length(df)
H <- 20
r <- model$fit$residuals[1:H]
acf.residuals <- acf(r, H, main = "ACF of residuals")
```

ACF of residuals



```
sum.denominator <- n - seq(H, 1, by = -1)
r.s <- acf.residuals$acf^2/sum.denominator
Q <- n * (n + 2) * sum(r.s)
Q
```

```
## [1] 720.2996
```

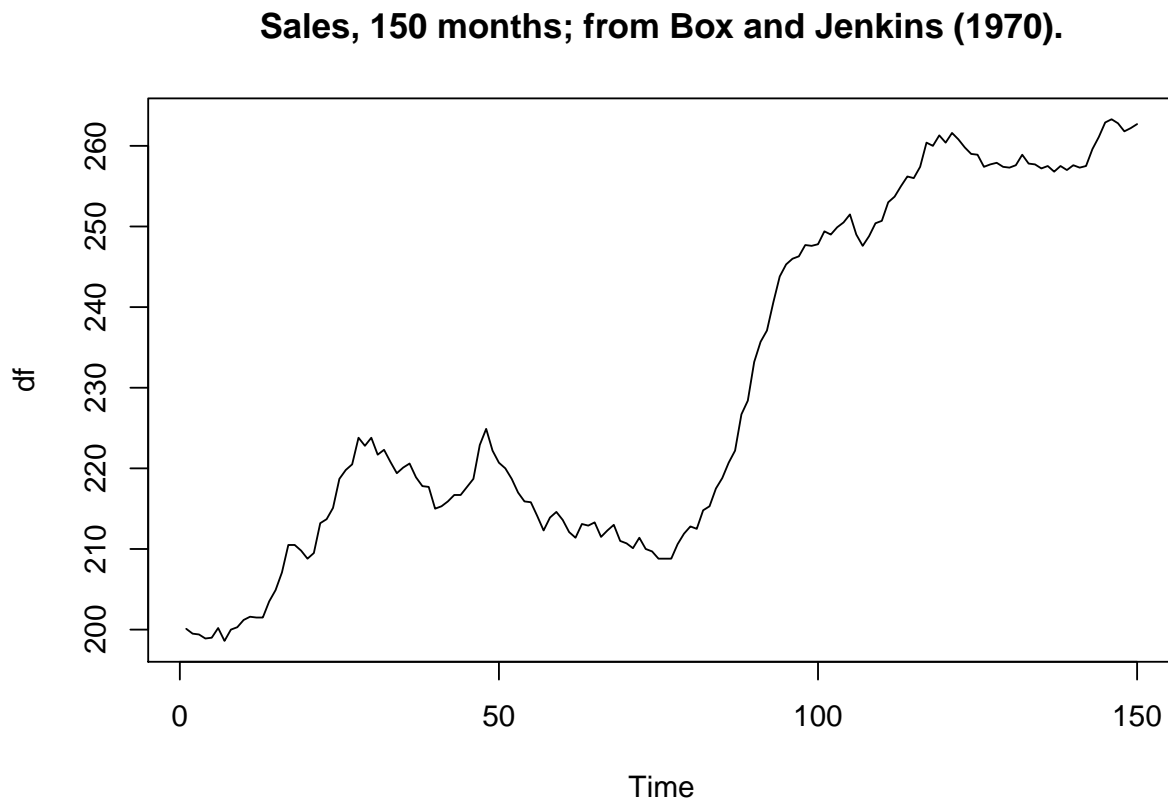
We see based on the Q-statistic that we have significant correlation structure remaining in the residuals.

We didn't expect the residuals to be normally distributed. Starting in 2005 there is a change in the volatility.

3.35 Sales analysis

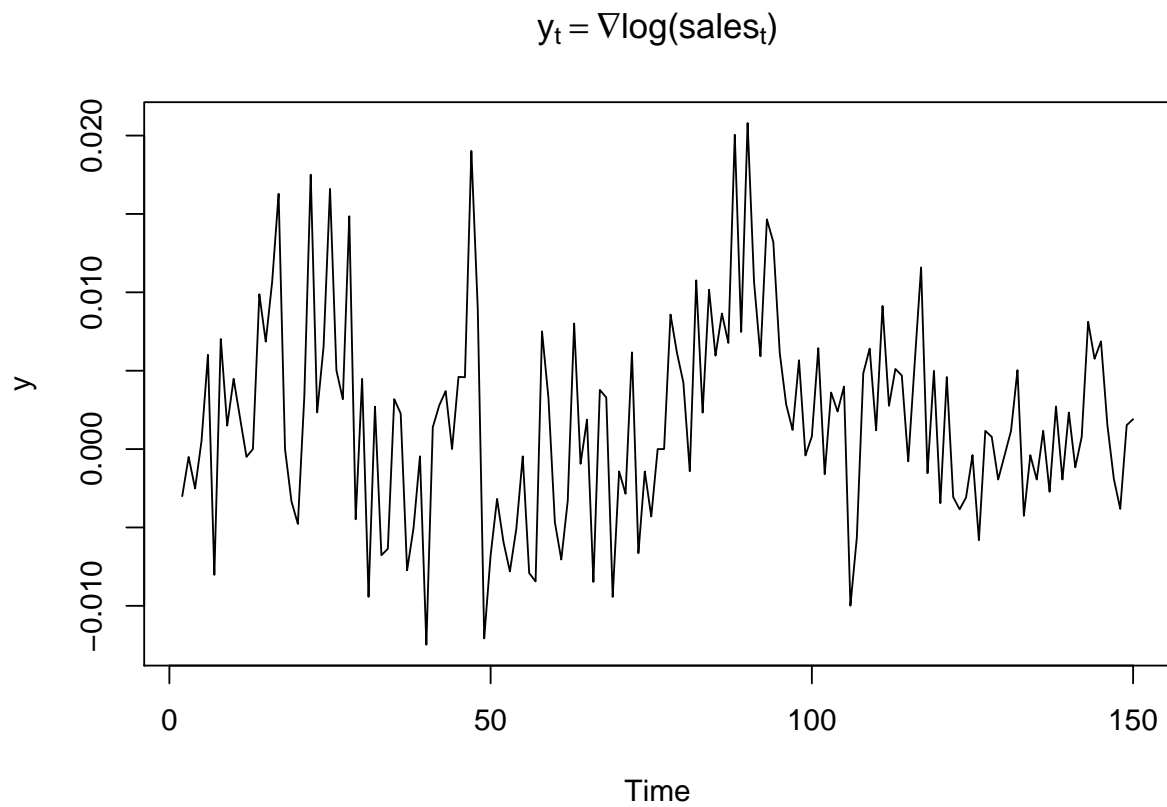
Let S_t represent the monthly sales data in sales ($n = 150$), and let L_t be the leading indicator in lead. Fit an ARIMA model to S_t , the monthly sales data. Discuss your model fitting in a step-by-step fashion, presenting your (A) initial examination of the data, (B) transformations, if necessary, (C) initial identification of the dependence orders and degree of differencing, (D) parameter estimation, (E) residual diagnostics and model choice

```
rm(list = ls())
data(sales, package = "astsa")
df <- sales
plot(df, main = "Sales, 150 months; from Box and Jenkins (1970).")
```



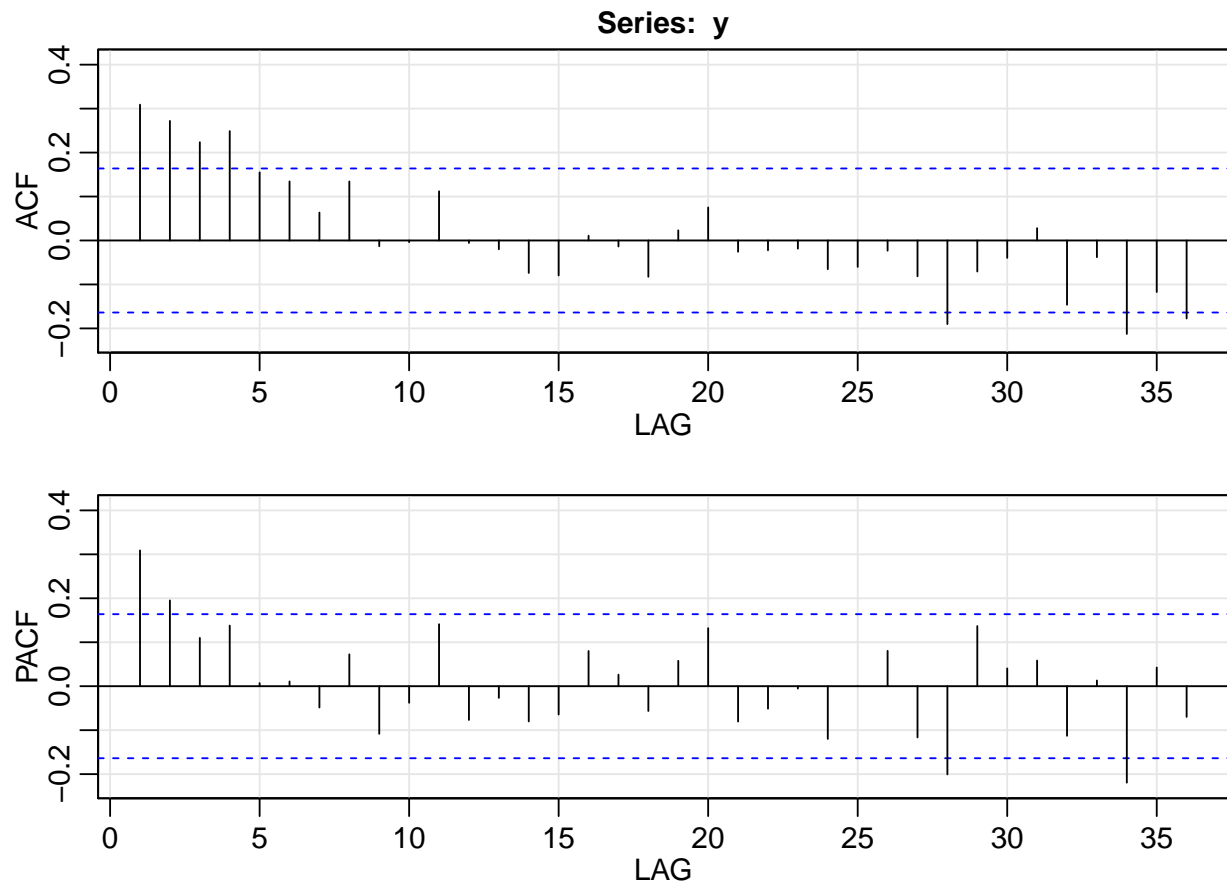
For our modelling we'll use a growth rate transformation.

```
y <- diff(log(df))
plot(y, main = TeX("$y_t = \\nabla \log(\text{sales}_t)$"))
```



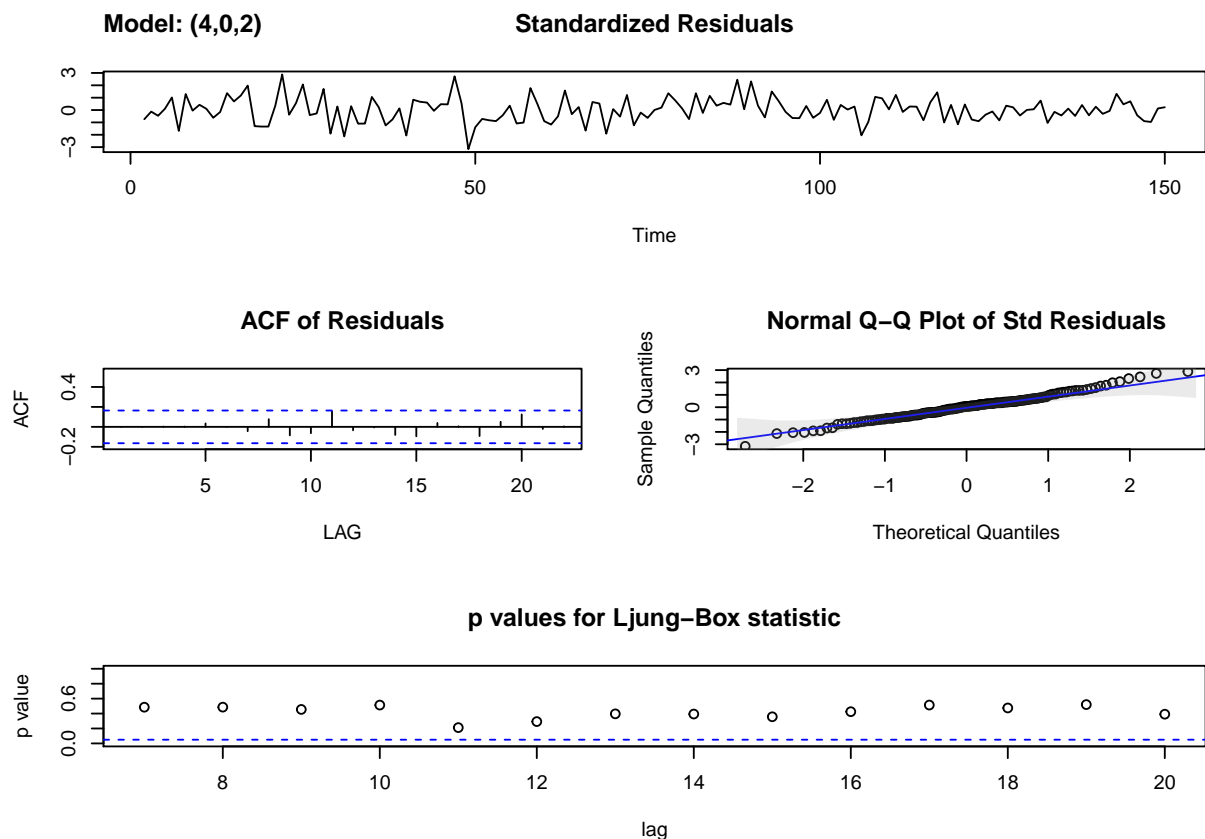
Now we'll investigate the ACF/PACF plots to determine appropriate (p, q) for our model.

```
acf2(y, 36)
```

Based on the ACF/PACF we will try to model $ARMA(4,2)$ first and consider models of lower order based on experimentation.

```
invisible(model <- sarima(y, 4, 0, 2))
```



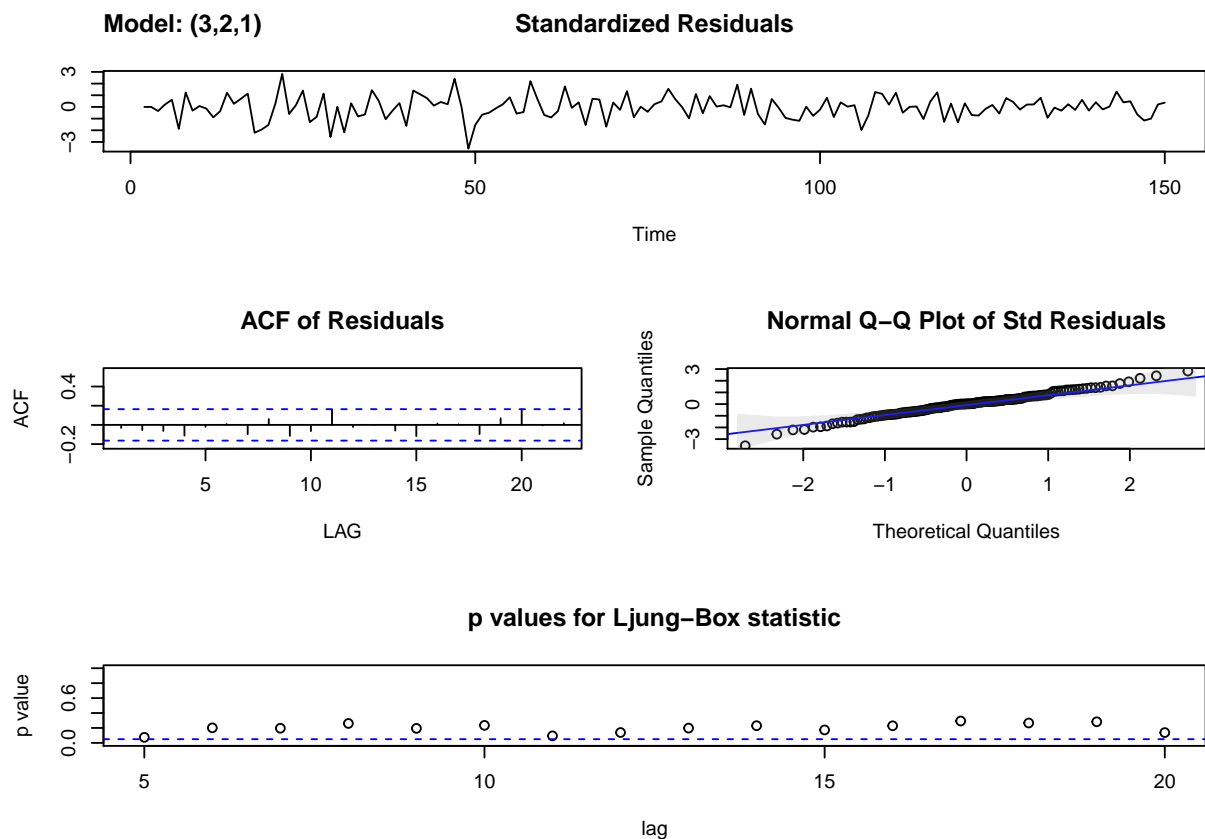
```
model
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##      Q), period = S), xreg = xmean, include.mean = FALSE, optim.control = list(trace =
##      REPORT = 1, reltol = tol))
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ma1      ma2      xmean
##    -0.0986  0.4502  0.0742  0.1235  0.3131 -0.2456  0.0017
## s.e.   0.4086  0.3129  0.1523  0.1092  0.4096   0.3170  0.0011
##
## sigma^2 estimated as 3.436e-05:  log likelihood = 554.18,  aic = -1092.37
##
## $degrees_of_freedom
## [1] 142
##
## $tttable
##      Estimate      SE t.value p.value
```

```
## ar1    -0.0986  0.4086 -0.2412  0.8097
## ar2     0.4502  0.3129  1.4385  0.1525
## ar3     0.0742  0.1523  0.4873  0.6268
## ar4     0.1235  0.1092  1.1310  0.2599
## ma1     0.3131  0.4096  0.7644  0.4459
## ma2    -0.2456  0.3170 -0.7747  0.4398
## xmean    0.0017  0.0011  1.5434  0.1250
##
## $AIC
## [1] -9.184628
##
## $AICc
## [1] -9.164302
##
## $BIC
## [1] -10.0435
```

This is a terrible model! There is still strong trend. We will try further differencing. We'll also experiment with the orders.

```
invisible(model <- sarima(y, 3, 2, 1))
```



```

model

## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##      Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc,
##      REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1          ar2          ar3          ma1
##      -0.6676  -0.4218  -0.2434  -1.0000
## s.e.    0.0799   0.0905   0.0794   0.0184
##
## sigma^2 estimated as 3.826e-05:  log likelihood = 535.36,  aic = -1060.72
##
## $degrees_of_freedom
## [1] 143
##
## $ttable
##      Estimate      SE  t.value p.value
## ar1  -0.6676 0.0799  -8.3554  0.0000
## ar2  -0.4218 0.0905  -4.6588  0.0000
## ar3  -0.2434 0.0794  -3.0670  0.0026
## ma1  -1.0000 0.0184 -54.4122  0.0000
##
## $AIC
## [1] -9.117437
##
## $AICc
## [1] -9.101198
##
## $BIC
## [1] -10.03679
T

```

3.43 Fit a seasonal ARIMA model of your choice to the U.S. Live Birth Series (birth).

Use the estimated model to forecast the next 12 months.

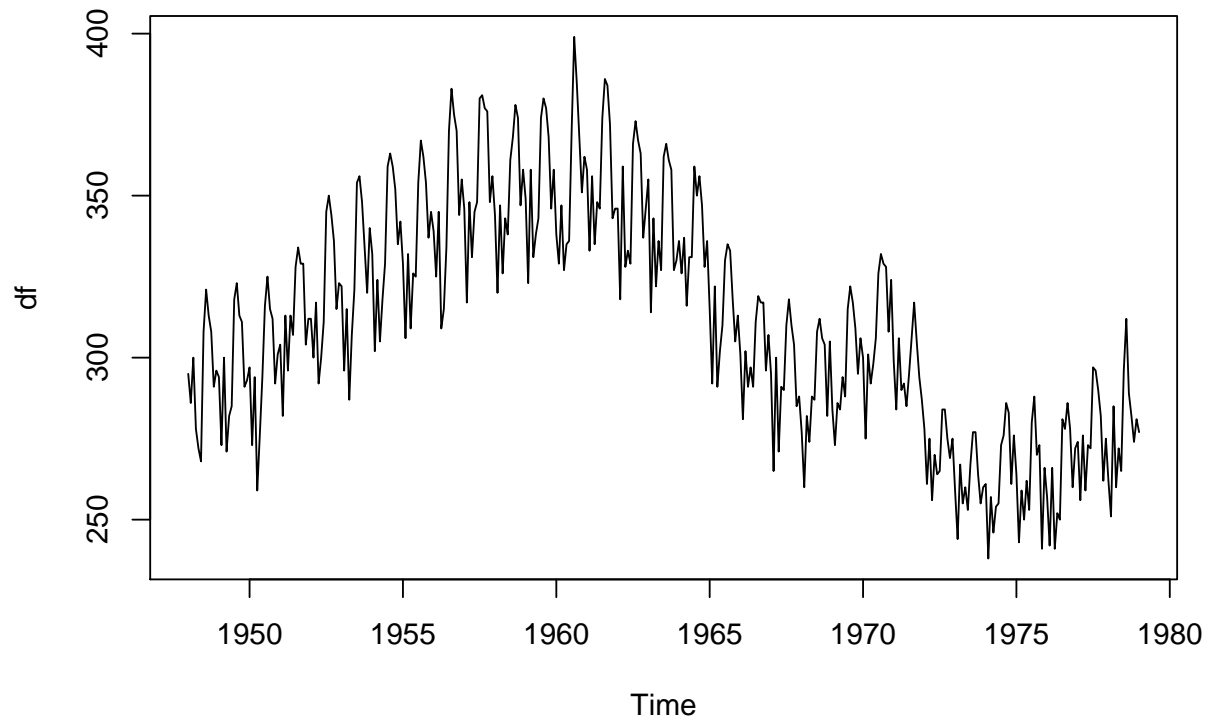
```

data(birth, package = "astsa")
df <- birth

```

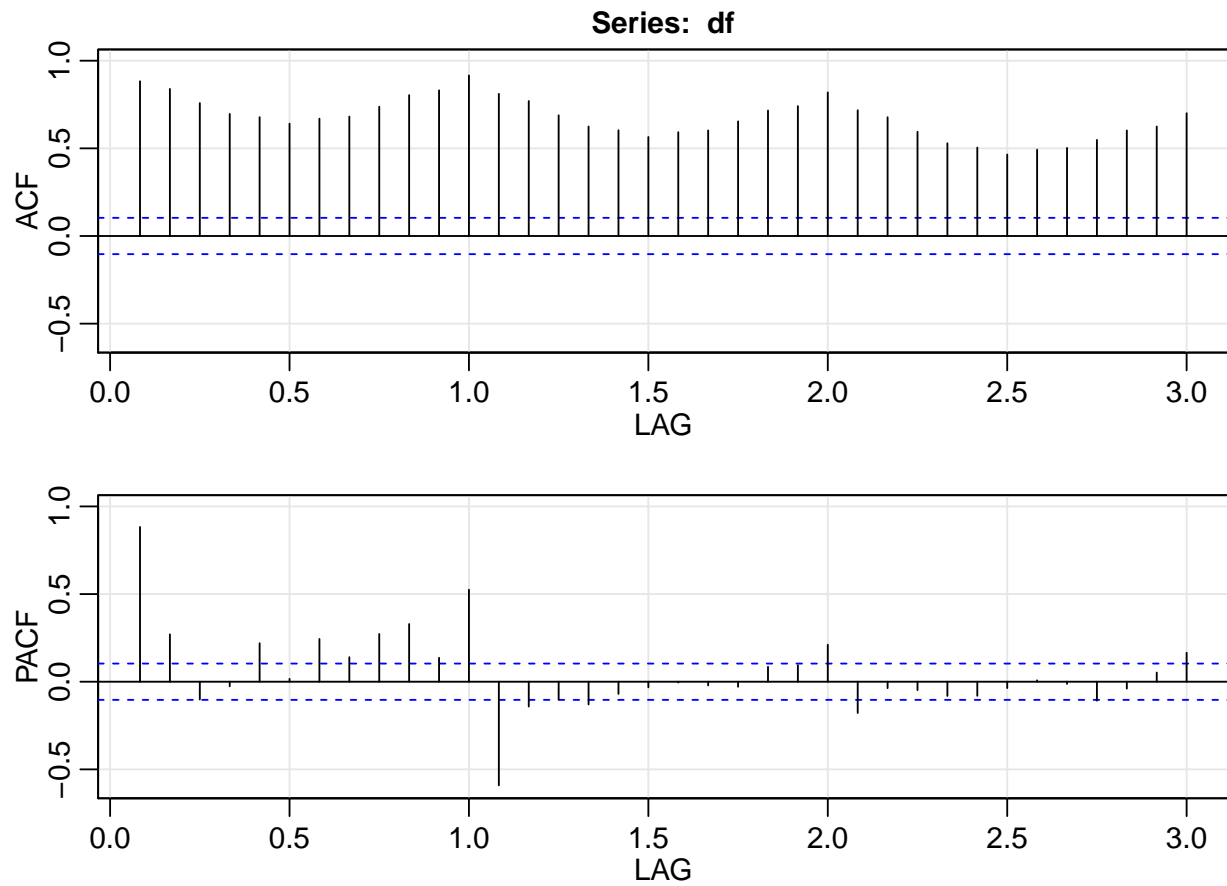
```
plot(df, main = "Monthly live births (adjusted) in thousands for the United States, 1948–1979")
```

Monthly live births (adjusted) in thousands for the United States, 1948–1979



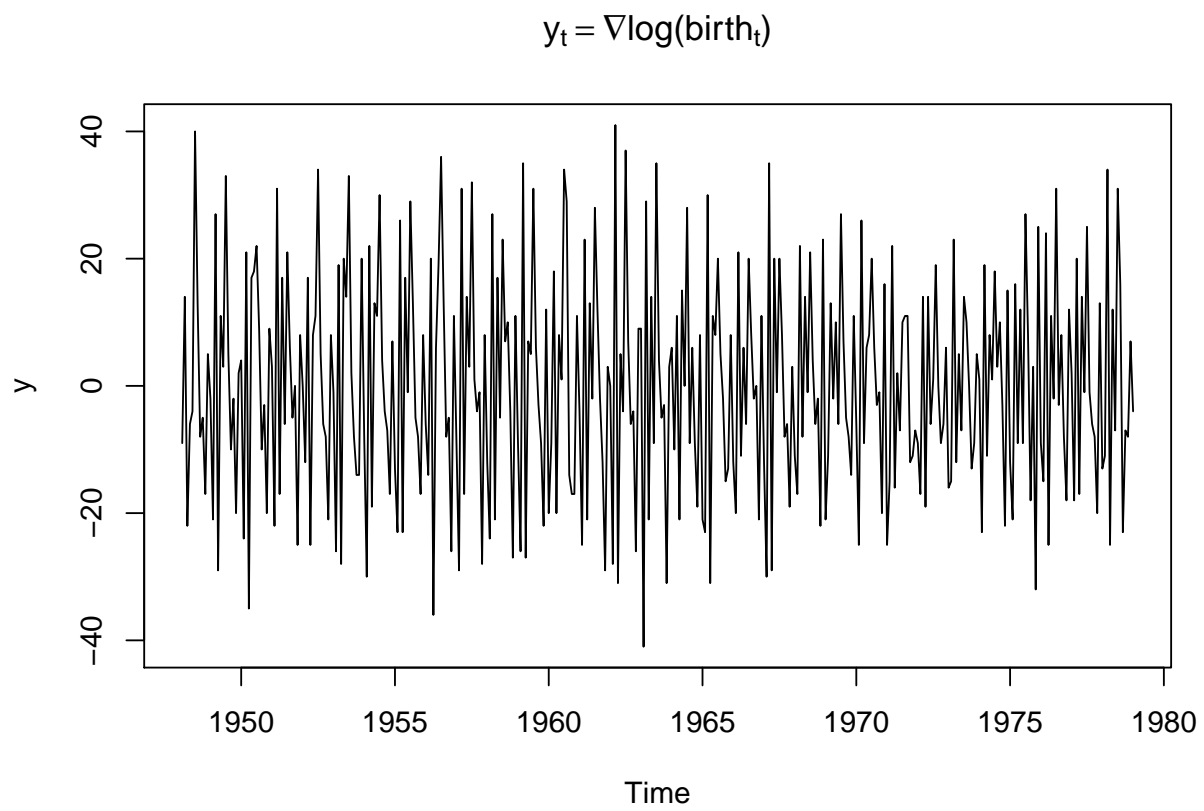
Here's the ACF / PACF of the series

```
acf2(df, 36)
```

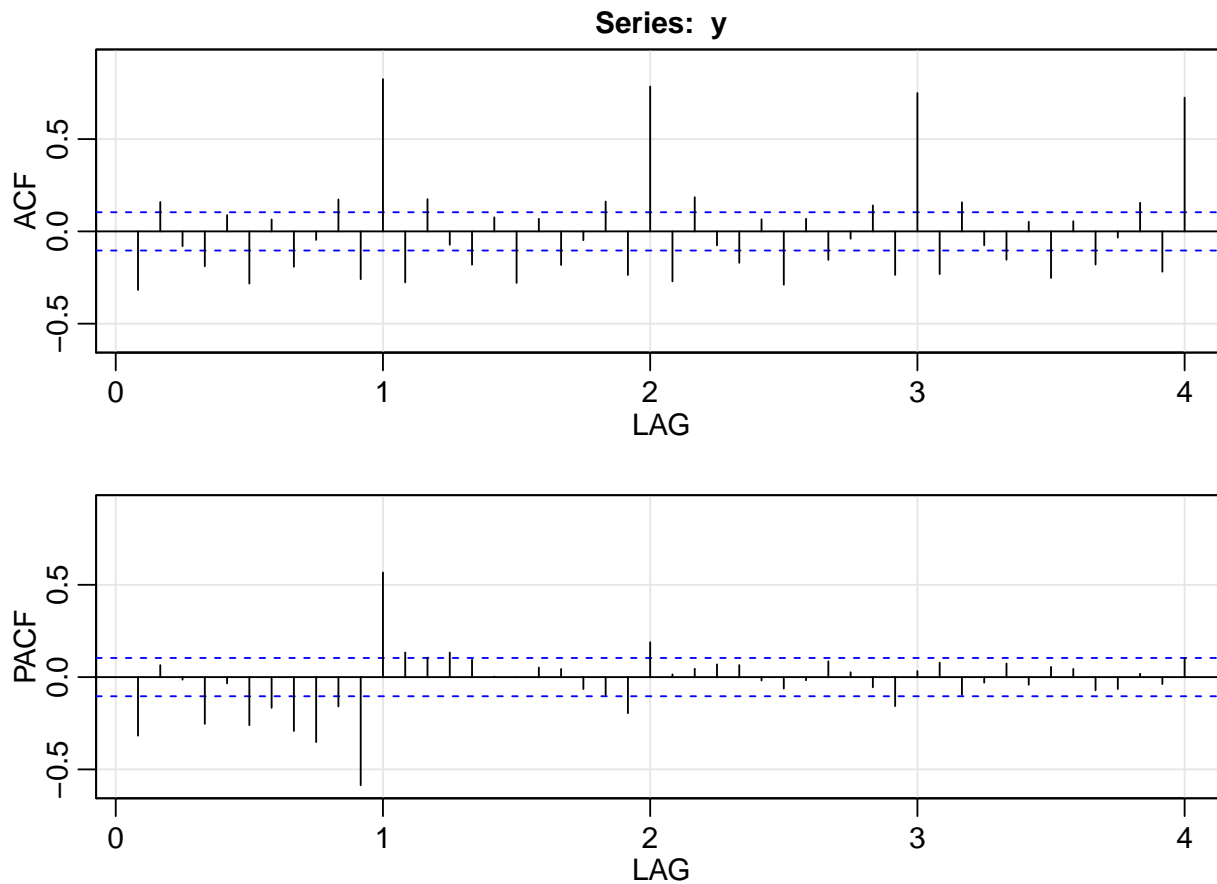


We see significant long range correlation so we look to differencing.

```
y <- diff((df))
plot(y, main = TeX("$y_t = \\nabla \log(\text{birth}_t)$"))
```



```
acf2(y, 48)
```



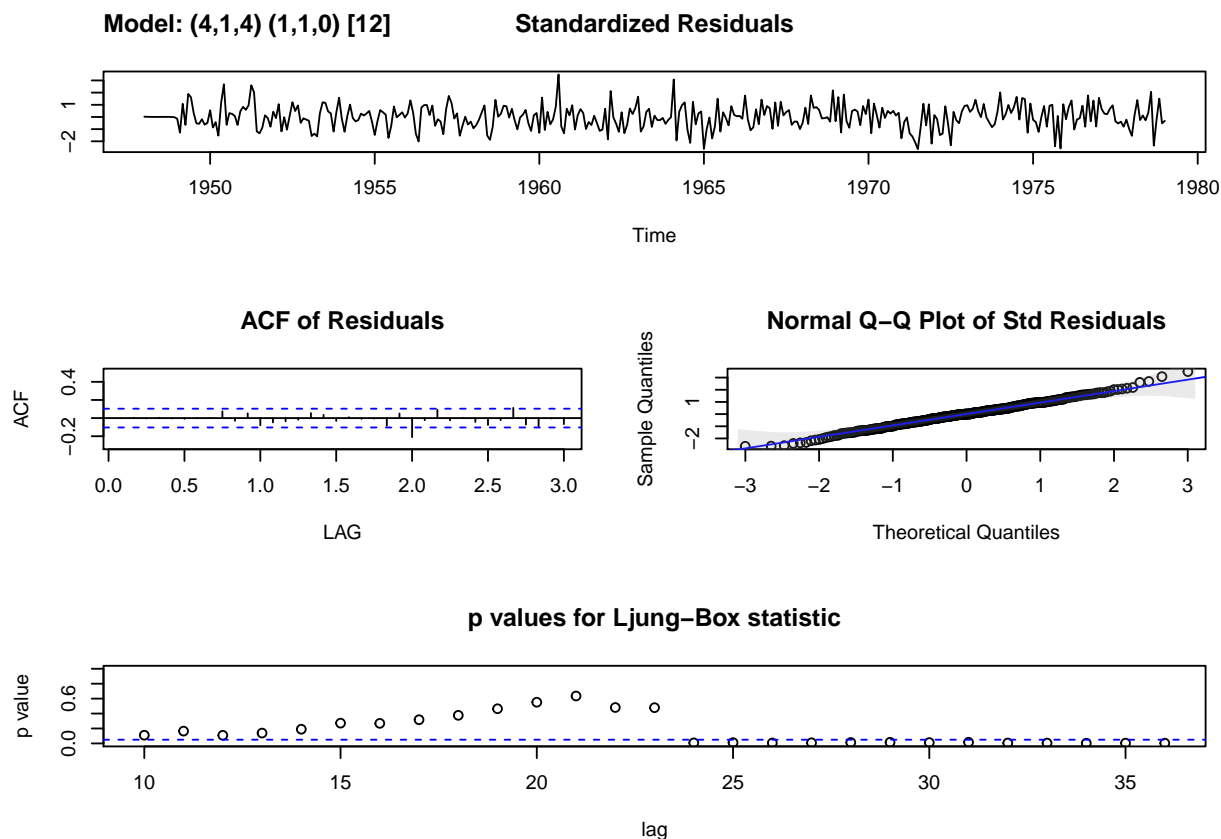
It's clear from this plot that we'll need a seasonal AR component with lag 12. It's not clear yet that we need to seasonally difference ∇_{12} so we'll fit multiple models and evaluate them by information theoretic criteria. Based on the ACF and PACF of the differenced series and some experimentation - we will try to fit these models

$$ARIMA(4, 1, 4) \times (1, 1, 0)_{12}$$

and

$$ARIMA(2, 1, 2) \times (1, 1, 1)_{12}$$

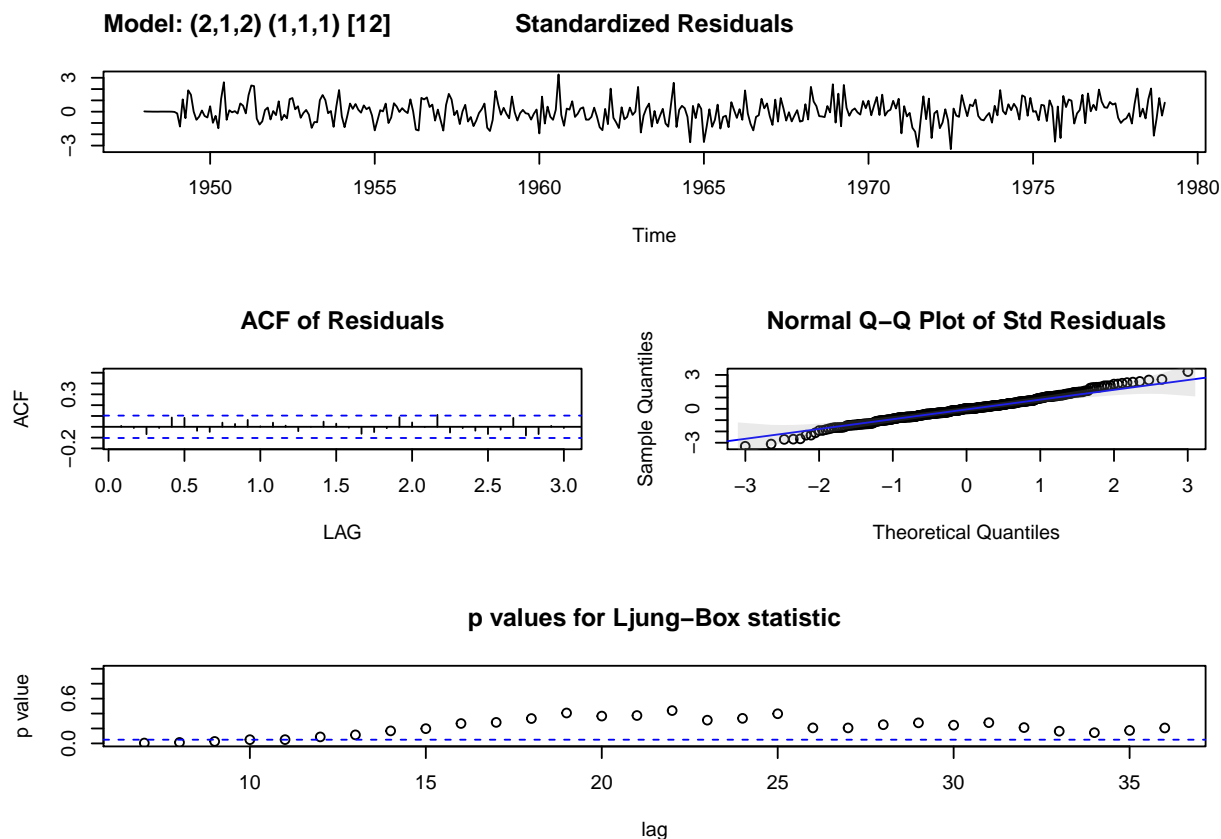
```
invisible(model <- sarima(df, 4, 1, 4, 1, 1, 0, 12))
```

```
model
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ma1      ma2      ma3      ma4
##      0.3230 -1.1199  0.2309 -0.2266 -0.7643  1.1612 -0.7456  0.1793
## s.e.  0.3721  0.1144  0.3360  0.0997  0.3768  0.2603  0.3713  0.2517
##      sar1
##      -0.4290
## s.e.   0.0491
##
## sigma^2 estimated as 52.06:  log likelihood = -1225.03,  aic = 2470.07
##
## $degrees_of_freedom
## [1] 351
```

```
##
## $ttable
##      Estimate      SE t.value p.value
## ar1      0.3230 0.3721  0.8681  0.3860
## ar2     -1.1199 0.1144 -9.7926  0.0000
## ar3      0.2309 0.3360  0.6873  0.4924
## ar4     -0.2266 0.0997 -2.2724  0.0237
## ma1     -0.7643 0.3768 -2.0285  0.0433
## ma2      1.1612 0.2603  4.4605  0.0000
## ma3     -0.7456 0.3713 -2.0080  0.0454
## ma4      0.1793 0.2517  0.7123  0.4767
## sar1    -0.4290 0.0491 -8.7312  0.0000
##
## $AIC
## [1] 5.000733
##
## $AICc
## [1] 5.007724
##
## $BIC
## [1] 4.095355
invisible(model <- sarima(df, 2, 1, 2, 1, 1, 1, 12))
```



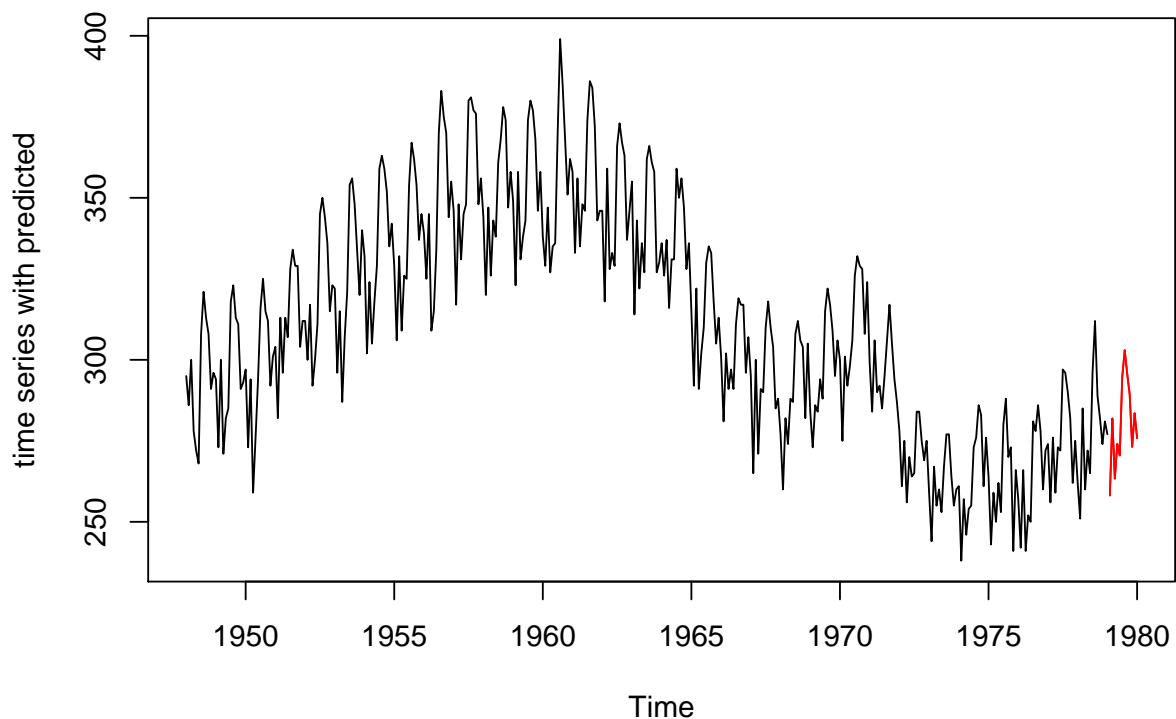
```
model
```

```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), include.mean = !no.constant, optim.control = list(trace = trc,
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ar2      ma1      ma2      sar1      sma1
##      -0.6109  0.2735  0.2178 -0.6506  0.1102 -0.8475
## s.e.   0.2118  0.1153  0.1997  0.1535  0.0722  0.0475
##
## sigma^2 estimated as 45.5:  log likelihood = -1204.71,  aic = 2423.43
##
## $degrees_of_freedom
## [1] 354
##
## $tttable
##      Estimate      SE  t.value p.value
```

```
## ar1    -0.6109  0.2118  -2.8843  0.0042
## ar2     0.2735  0.1153   2.3730  0.0182
## ma1     0.2178  0.1997   1.0907  0.2762
## ma2    -0.6506  0.1535  -4.2382  0.0000
## sar1     0.1102  0.0722   1.5266  0.1278
## sma1    -0.8475  0.0475 -17.8599  0.0000
##
## $AIC
## [1] 4.849796
##
## $AICc
## [1] 4.85598
##
## $BIC
## [1] 3.912878
```

Based on the AIC criterion We choose to use the $ARIMA(2,1,2) \times (1,1,1)_{12}$ model for prediction.

```
fore = predict(model$fit, n.ahead = 12, interval = "prediction")
ts.plot(df, fore$pred, col = 1:2, ylab = "time series with predicted")
lines(fore$pred, pch = "*", col = 2)
```



```
# lines(fore$pred+1.96*fore$se, lty='dashed', col=4)
# lines(fore$pred-1.96*fore$se, lty='dashed', col=4)
```

Our model worked as expected.

WRONG PROBLEM THIS IS FROM VERISON 3 OF THE TEXTBOOK 3.35 Seasonal Model

Consider the ARIMA model $x_t = w_t + \Theta w_{t-2}$.

(a) Identify the model using the notation $ARIMA(p, d, q)$ or $(P, D, Q)_s$

This model is $ARIMA(0, 0, 1) \times (0, 0, 1)_2$ It's worth noting that it's also a $MA(2)$ with $\theta_1 = 0$

(b) Show that the series is invertable for $|\Theta| < 1$, and find the coefficients in the representation $w_t = \sum_{k=0}^{\infty} \pi_k x_{t-k}$.

$\phi(z) = 1 - \Theta z^2$ If $|\Theta| < 1$ the roots of $\phi(z)$ $z_o = \pm \frac{1}{\sqrt{\Theta}}$ are outside the unit circle and we have that the series is invertible.

WRONG - AND VERY DIFFICULT - PROBLEM. THIS IS FROM VERISON 3 OF THE TEXTBOOK 3.43

Use Theorem B.2 and B.3 to verify (3.116).

I actually worked on this for a day or so. If I have time, I'd like to revisit the appendix and this problem. The problem asks us to validate the likelihood formula for a normal ARMA model using the projection operator and a property of conditional expectation in Hilbert space. Here's what I came up with so far.

We know that the MLE is provided by OLS, and that the projection given by the hat matrix provides the OLS transform. This is the projection on the span of the $\{1, z_j\}$ from Theorem B.3.

Theorem B.2 provides a mechanism to build up estimates through projections onto nested subspaces. It's important to note that when we're doing ML we take the data to be random variables and ask what values of the parameters best explain the data at hand. The Hilbert space we use is formed from the space of random variables X_t . Since the process is second order stationary we know that all the information of the process is contained in the second order statistics. The Hilbert space inner product is the covariance.

The expression for the likelihood in 3.116 looks like a transformation of the likelihood expressed by the full joint distribution to the likelihood expressed as a product of conditional distributions - but there is a lot more to it than that! This is a complicated expression

based on the past section where we calculated the MLE of x_t^{t-1} . We express the conditional probabilities in terms of x_t^{t-1} , $(x_t|x_{t-1}, \dots, x_1) \sim N(x_t^{t-1}, P_t^{t-1})$

Since we know that OLS is a projection, and that the MLE is provided by OLS, one can see that the process of maximizing 3.116 can be cast as a problems of forming the right sequence of nested subspaces and applying the projection sequentially. The sequence of spaces is

$$\text{span}\{x_1\} \subset \text{span}\{x_1, x_2\} \subset \dots \subset \text{span}\{x_1, \dots, x_t\}$$

The hat matrix formed in each of these subspaces provides the projection we need to maximize the corresponding factor in the likelihood 3.116.

What I don't understand yet is how to use the projection theorems to verify the likelihood expression 3.116.