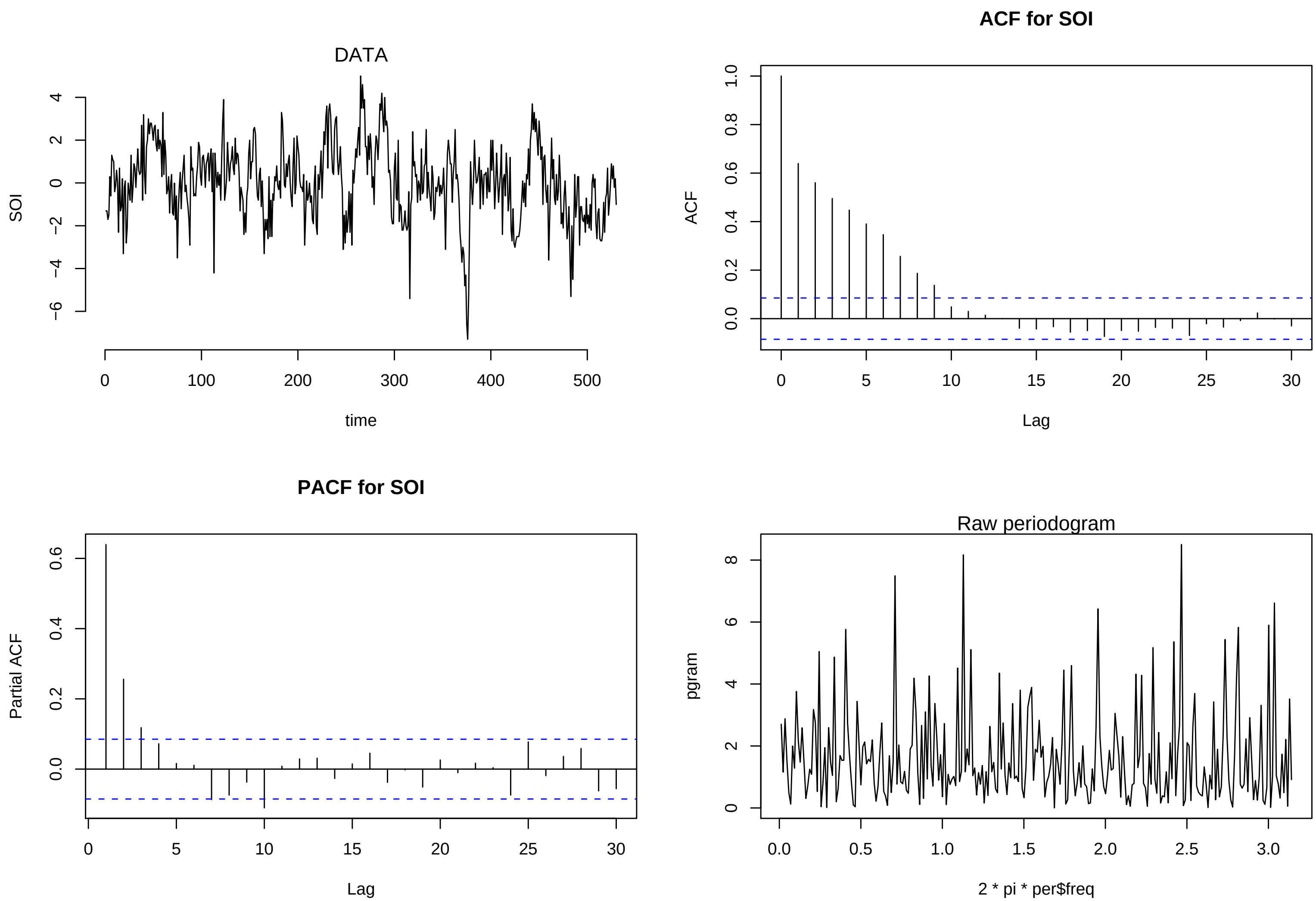
SOI data example. 540 monthly values, 1950-1995

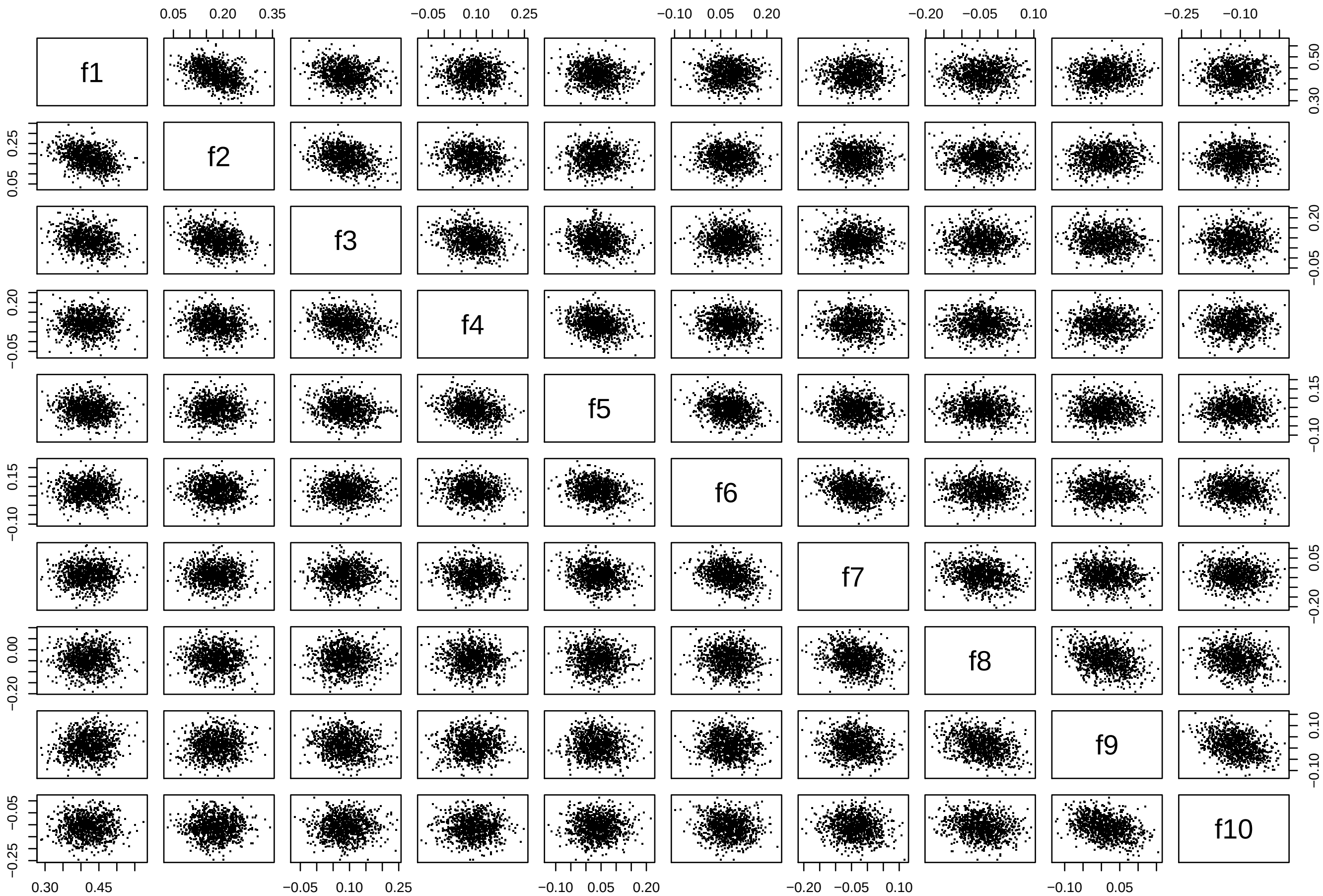
- This series is computed as the "difference of the departure from the long-term monthly mean sea level pressures" at Tahiti in the South Pacific and Darwin in Northern Australia.
- The index is one measure of the so-called "El Nino-Southern Oscillation".
- The fact that most of the observations in the last part of the series take negative values has been related to a recent warming in the tropical Pacific.
- A key question is to determine how unusual the event is, and if it can reasonably be explained by standard "stationary" time series models.


DATA
SOI
time
ACF for SOI
ACF
Lag
PACF for SOI
Partial ACF
Lag
Raw periodogram
pgram
2 * pi * per$freq

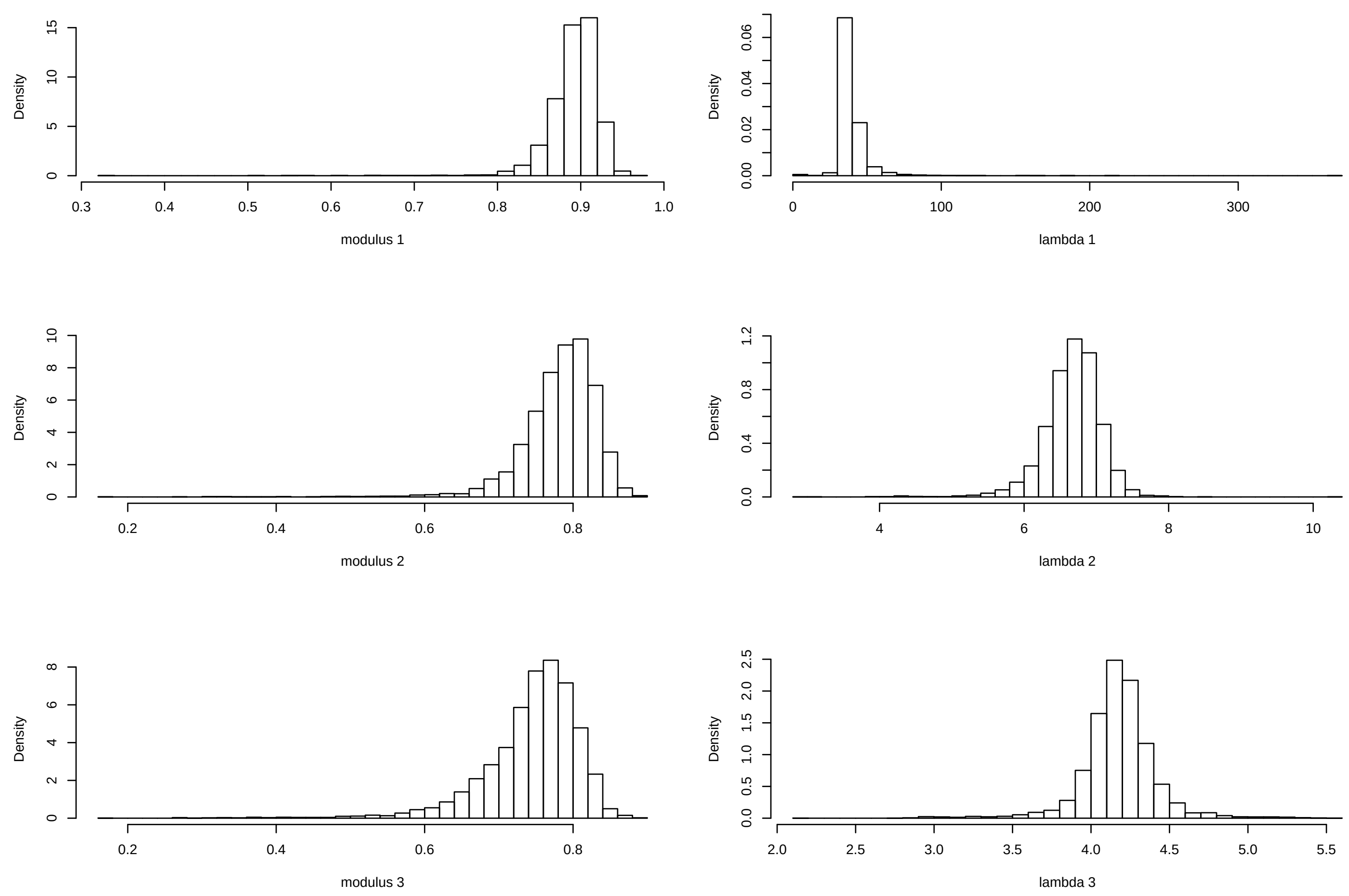
Results for “ar” function

```
a$order [1] 10
round(a$ar,digits=2)
 [1] 0.42 0.18 0.09 0.09 0.04 0.07 -0.03 -0.05
0.01 -0.11
 a$order.max 27
a$var.pred   1.65
 m[order(w)]  0.89 0.78 0.74 0.79 0.78
 w[order(w)]  0.17 0.94 1.50 2.14 2.84
 2*pi/w[order(w)] 36.74  6.67  4.18  2.93  2.21
```

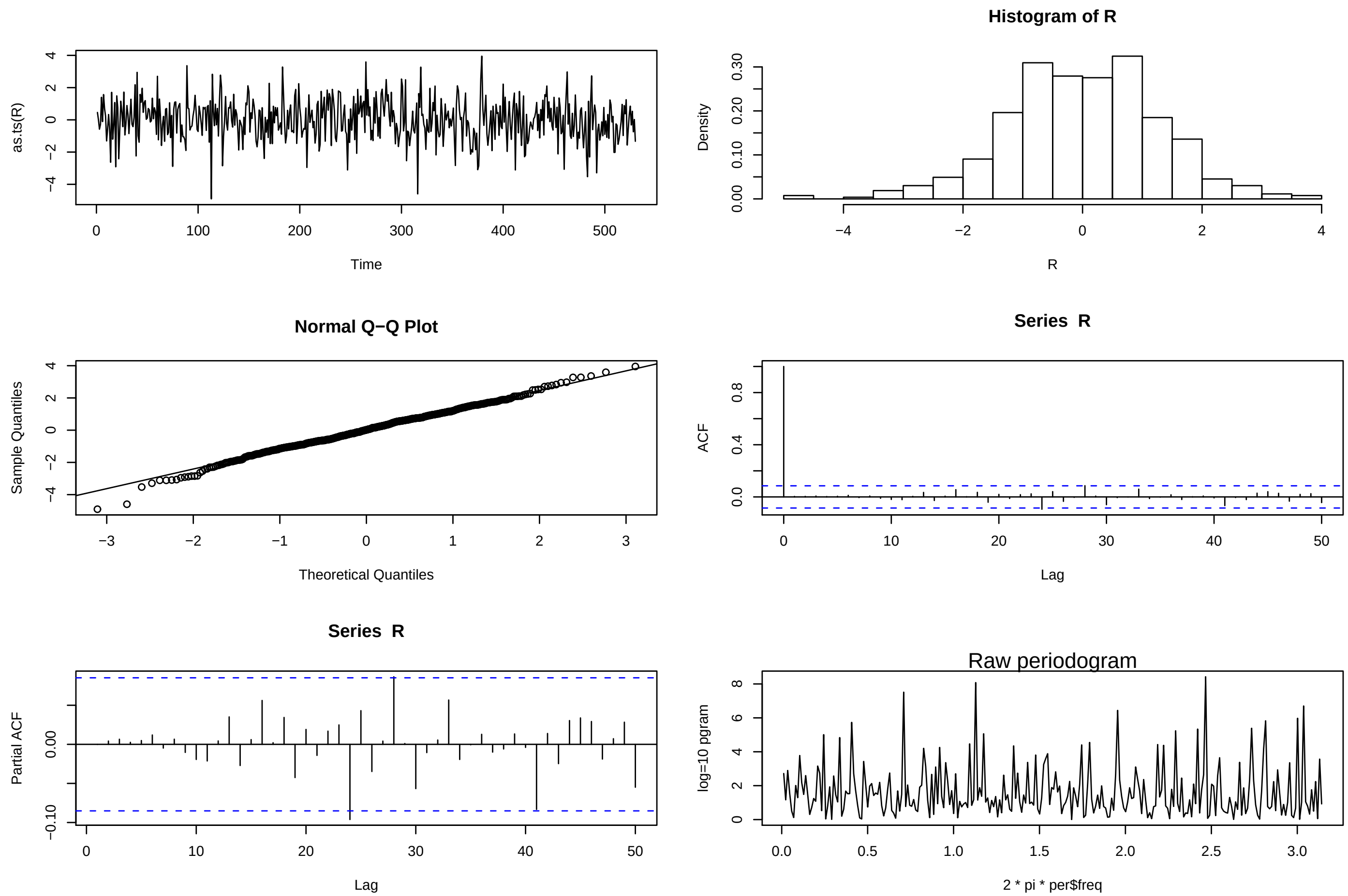
Plot by pairs of samples for $\phi_1 - \phi_{10}$



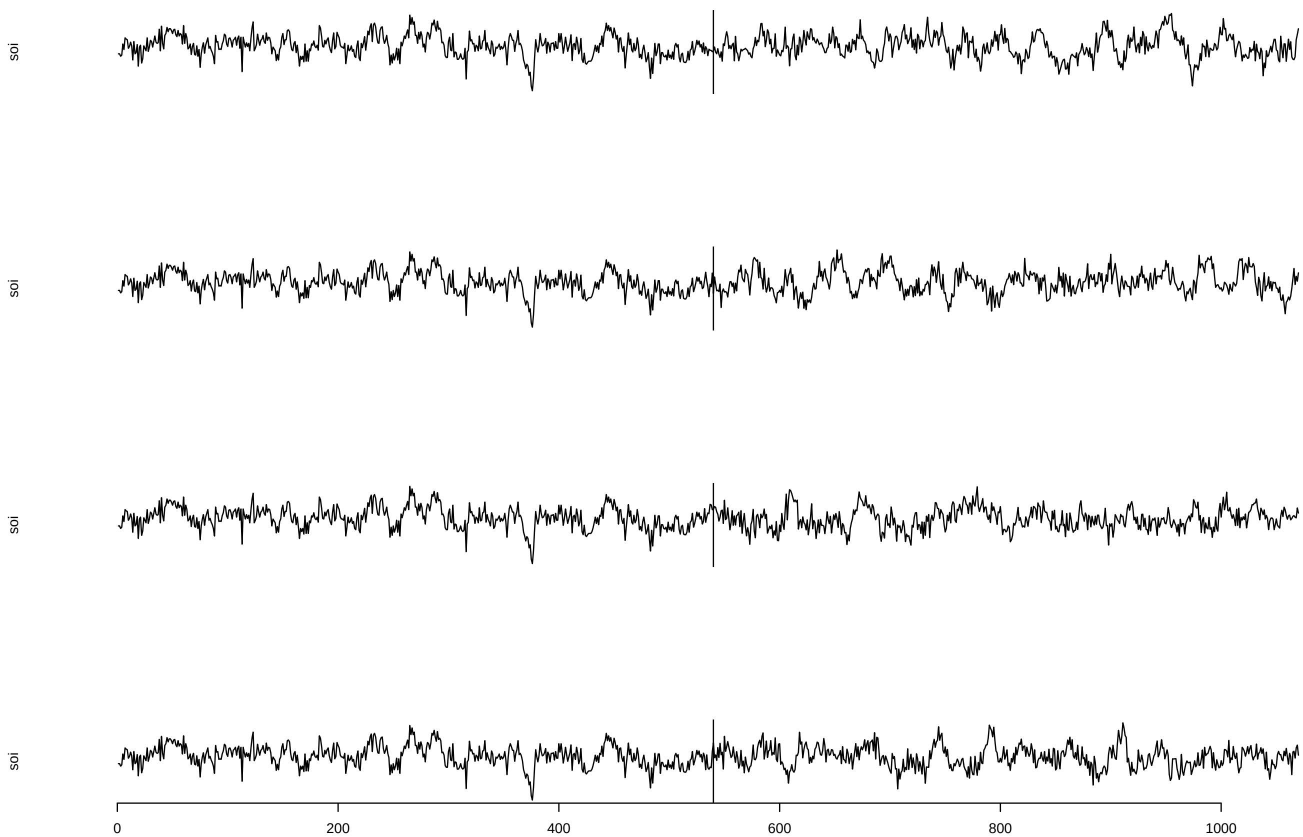
Posterior samples of reciprocal roots (r_i,λ_i)



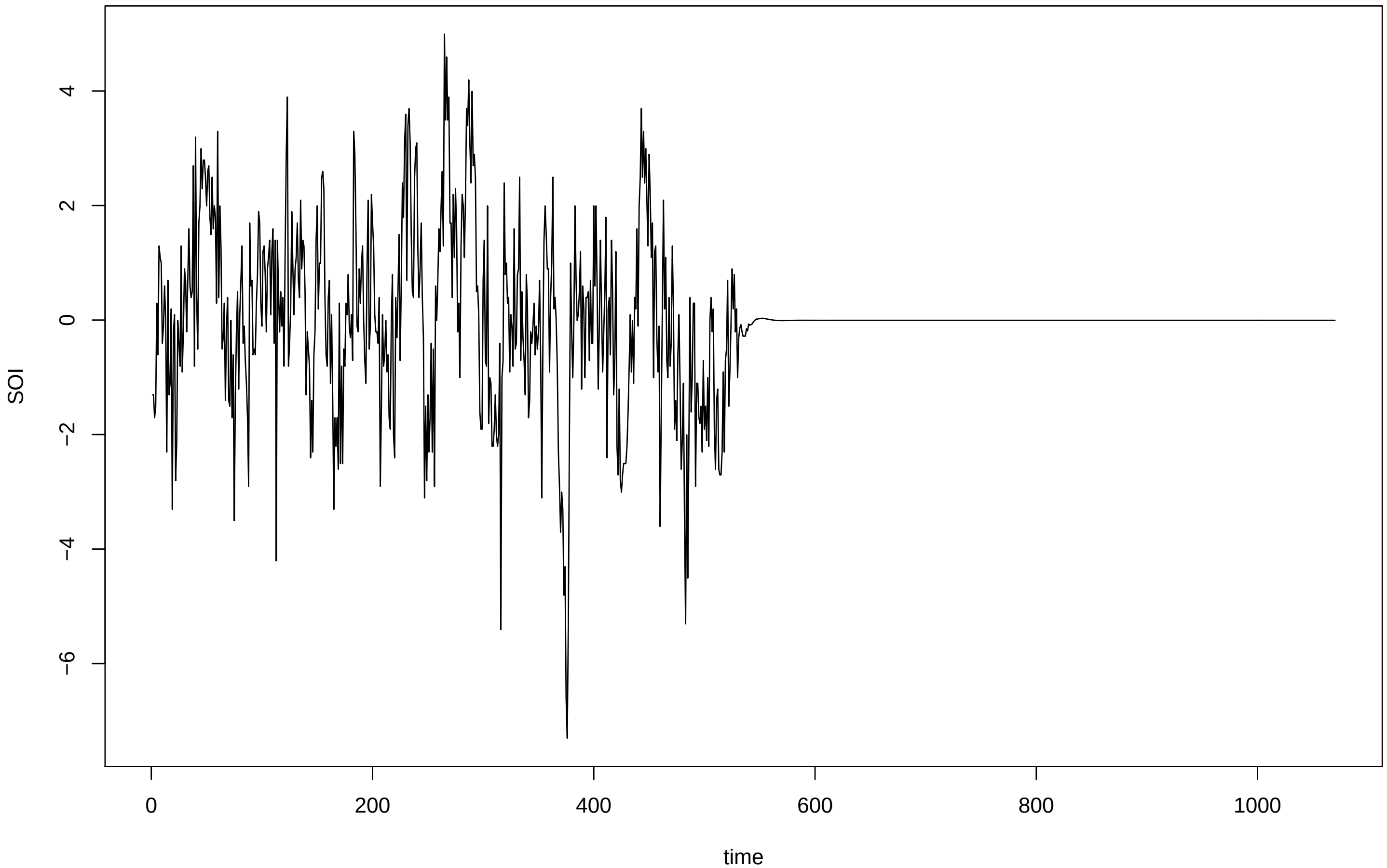
Plots based on residuals



SOI data and 4 draws from the predictive dist.



SOI data and MSE forecasts



Estimation for ARMA models

- Recall than an ARMA(p,q) is defined by the expression

$$\Phi(B)X_t = \Theta(B)\epsilon_t$$

or equivalently as

$$\epsilon_t = \theta_1\epsilon_{t-1} + \ldots + \theta_q\epsilon_{t-q} + X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p}$$

- Under the assumption of Normality and independence for the errors, ($\epsilon_t \sim N(0,\sigma^2)$) the log-likelihood of the ARMA model is

$$l(\phi,\theta,\sigma^2) = -(n/2)log(2\pi\sigma^2) - \sum_{t=1}^{n}\epsilon_t^2$$

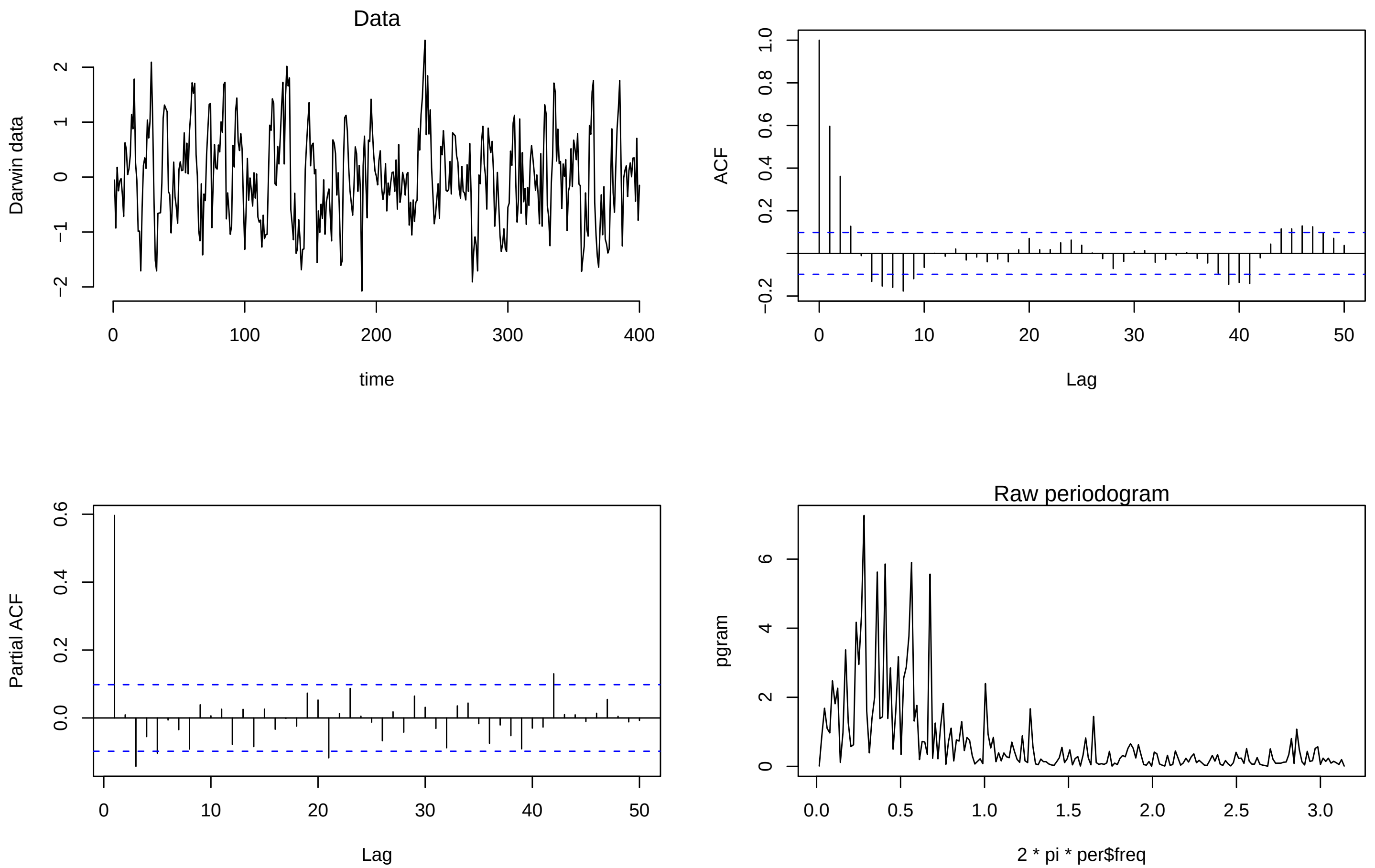
- Notice that ϵ_t depends on $\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q$ in a complicated way.

- To find the MLE for the model parameters numerical optimization is required.
- In R we can use the function *arima* to fit an ARMA to data.

```
arima(x,order=c(1,1))
# Fits an ARMA(1,1) to x
arima(x,order=c(1,1,1))
# Fits an ARIMA(1,1,1) to x
```

- As with the AR model there is some form of conditioning to initial values.
- If $m = max\{p, q\}$, it is assumed that the initial data values $x_1, x_2, \ldots x_m$ and the initial errors $\epsilon_1, \epsilon_2, \ldots \epsilon_m$ are completely known.

- Usually these error are set equal to zero, i.e. $\epsilon_t = 0, t = 1, 2, \ldots, m$
- As an example, we consider a smooth version of the Darwin sea level pressure series (data file "darwin2").
- The annual cycle was removed and each data point is average over adjacent 3 month periods MAM, JJA, SON, DJF.
- In a paper by Trenberth and Hoar (1996) The 1990-1995 El Niño-Southern oscillation event: Longest in Record. *Geophysical Research Letters*, **23**, 57-60.), an ARMA(3,1) model was considered for this data.


Data
Darwin data
time
2
1
0
−1
−2
0
100
200
300
400
ACF
Lag
1.0
0.8
0.6
0.4
0.2
−0.2
0
10
20
30
40
50
Partial ACF
Lag
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Raw periodogram
pgram
2 * pi * per$freq
6
4
2
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

```
x=scan(``darwin2'')
x=x-mean(x)
arma31=arima(x,order=c(3,0,1),include.mean=F)
names(arma31)
"coef" "sigma2"    "var.coef"
"loglik" "aic" "residuals"
# MLE results
round(arma31$coef,digits=2)
    ar1   ar2   ar3    ma1
    1.28 -0.32 -0.14 -0.71
 se 0.16  0.13  0.06  0.16

round(arma31$sigma2,digits=2)
 0.43
```

Method of Moments

- For a method of moments approach consider the MA(1) model $X_t = \epsilon_t - \theta\epsilon_{t-1}$.
- From previous calculations for the ACF, we have that $\rho_1 = -\theta/(1+\theta^2)$.
- If we substitute $\hat{\rho_1}$, for ρ_1, and solve the quadratic equation for θ, we get that the MOM estimator is

$$\hat{\theta} = \left(-1 \pm \sqrt{1 - 4\hat{\rho_1^2}}\right)/2\hat{\rho_1}$$

- For $\hat{\rho_1} = 0.5$ we have a unique solution $\hat{\theta} = -1/2\hat{\rho_1}$
- However there is no real solution if $|\hat{\rho_1}| > 0.5$.
- There are two solutions if $|\hat{\rho_1}| < 0.5$.

- In general, the MOM estimator for ARMA models with $q \geq 1$ is not used.
- Given that the model equation of an ARMA process depends on both the $X_t's$ and the lagged $\epsilon_t's$, we cannot use our multiple linear regression framework for Bayesian analysis.

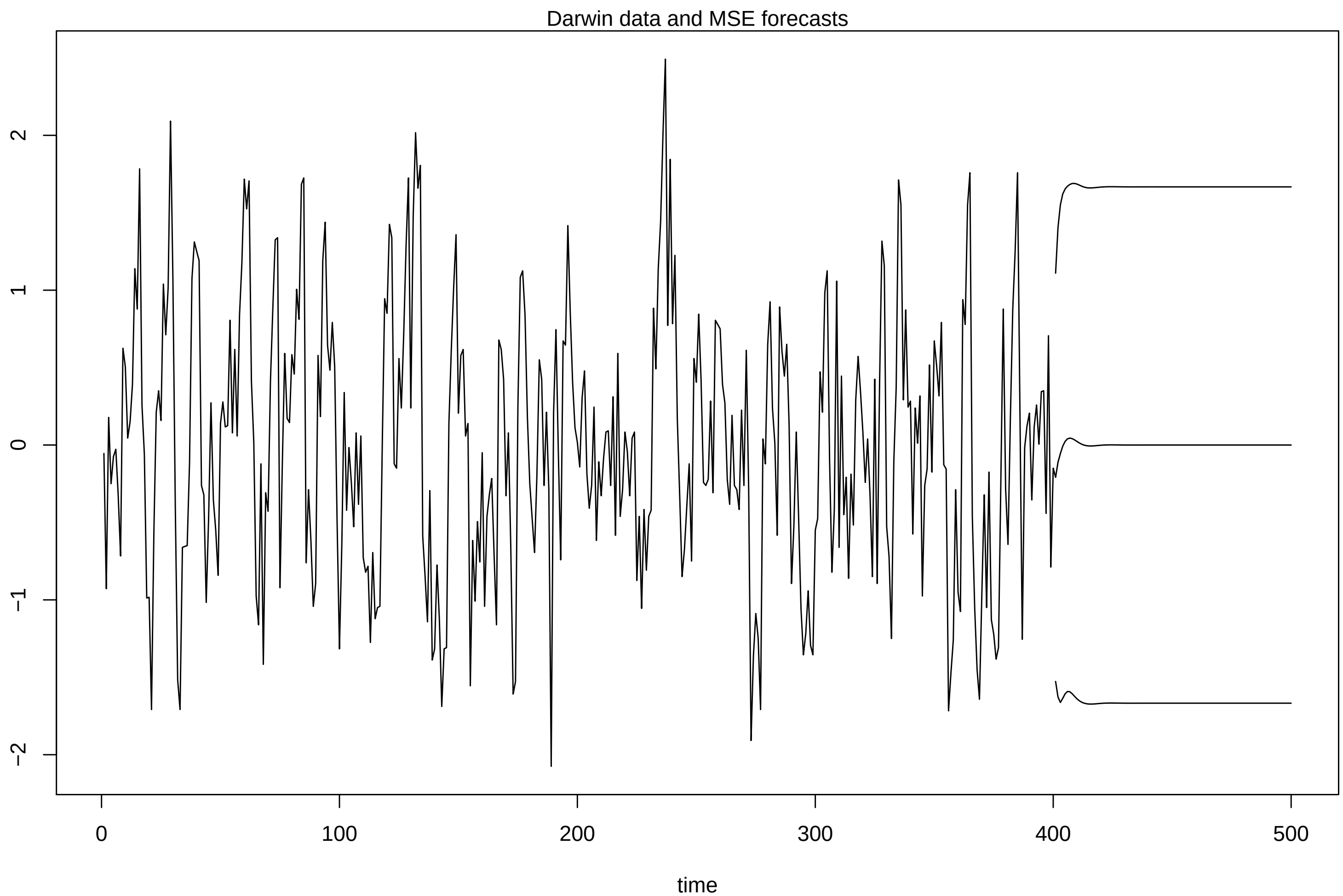
Forecasting

- Forecasting is similar to the AR case.
- We consider the infinite MA representation for the ARMA model,

$$X_t = (\phi(B))^{-1}\theta(B)\epsilon_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}$$

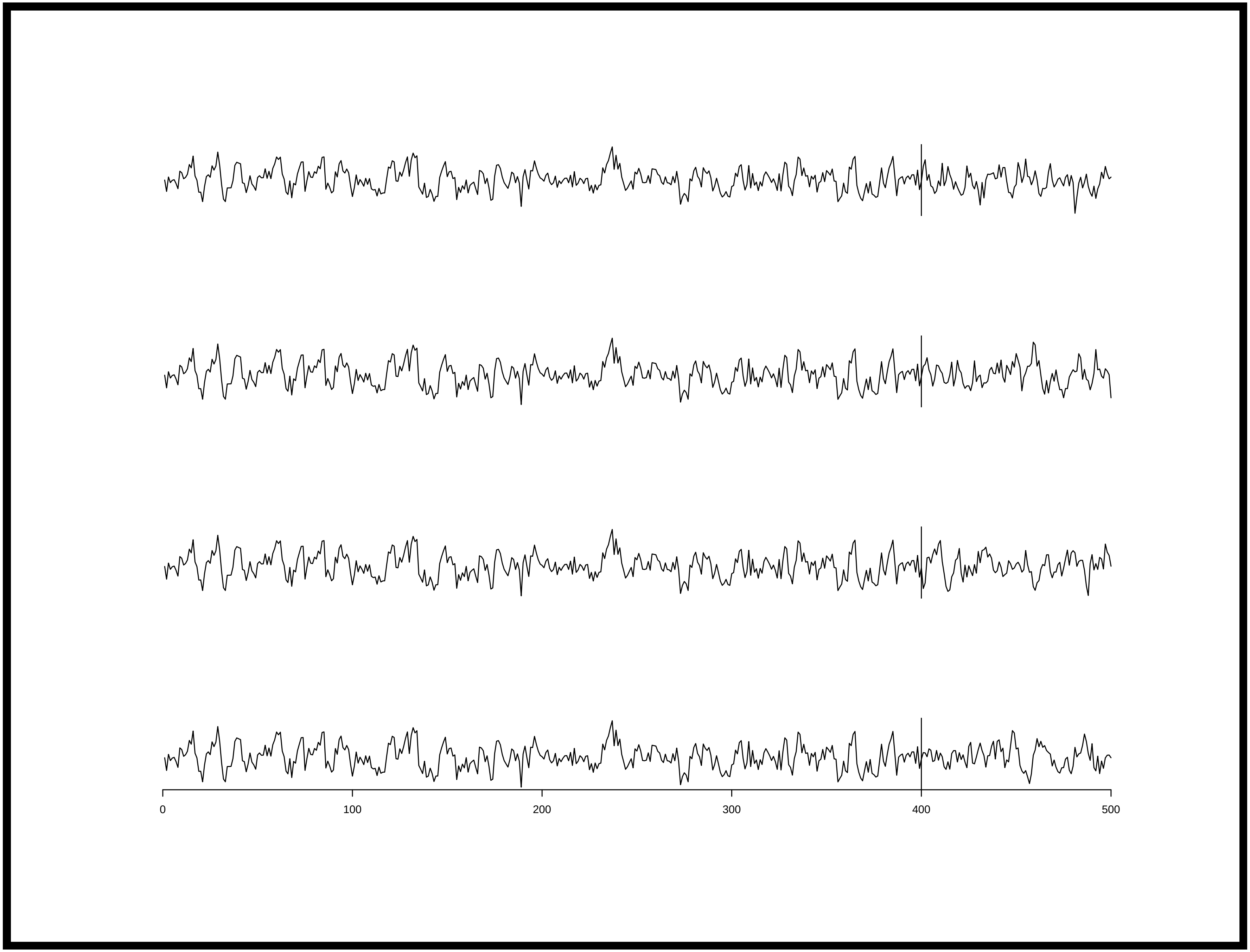
- It can be shown that the optimal MSE forecast for X_{t+k} is $E(X_{t+k}|X_n, X_{n-1}, \ldots)$, the conditional expectation of X_{t+k} given the data.
- Using R, for the Darwin data example the MSE forecasts for the next 100 time periods are computed as

```
darwin.fr=predict(arma31,n.ahead=100)
plot(c(x,darwin.fr$pred),type='l',xlab='time')
lines(401:500,darwin.fr$pred+2*darwin.fr$se)
lines(401:500,darwin.fr$pred-2*darwin.fr$se)
```


Darwin data and MSE forecasts
2
1
0
−1
−2
0
100
200
300
400
500
time

- We can simulate values of the process conditional on the MLE using the function *arima.sim* and treat them as a sample of “future” values.
- Formally this does not produce samples of the predictive distribution.

```
err=rnorm(400,mean=0,sd=sqrt(arma31$sigma2))
arcf=arma31$coef[1:3]
macf=arma31$coef[4]
sim=arima.sim(400,model=list(ar=arcf,ma=macf),innov=err)
sim=matrix(sim,4,100,byrow=T)
par(mfrow=c(4,1))
for(i in 1:4){plot(c(x,sim[i,]),axes=F,type='l',
xlab=' ',ylab=' ');abline(v=400)}
axis(1)
```


0
100
200
300
400
500

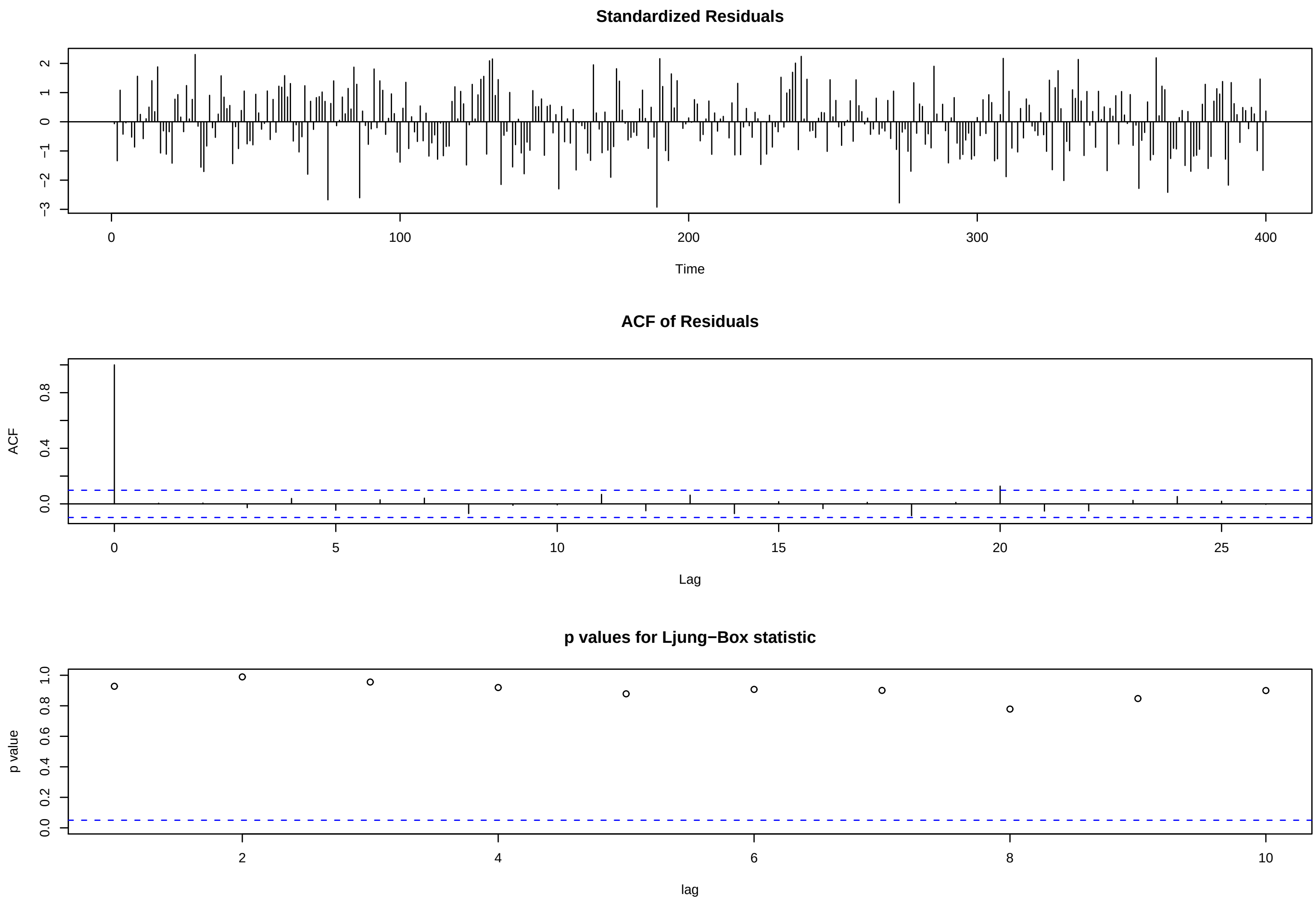
Diagnostics

- The main difference with respect to the AR analysis is the computation of the residuals.
- Conditional on initial values, the residuals are obtained recursively from the expression

$$\epsilon_t = \hat{\theta}_1 \epsilon_{t-1} + \ldots + \hat{\theta}_q \epsilon_{t-q} + X_t - \hat{\phi}_1 X_{t-1} - \ldots - \hat{\phi}_p X_{t-p}$$

```
#Residuals for Darwin example
arma31$residuals; tsdiag(arma31)
```

- A white noise test can be applied to these residuals or the Portmanteau (Box-Liung) test.
- The degrees of freedom for the chi-square test statistic are $K - (p + q)$ instead of $K - p$.


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung−Box statistic
p value
lag

Selection of p and q

- We can use AIC and BIC as with AR models. For ARMA models we a have functions of two arguments p and q.

$$\begin{aligned} AIC &= -2log(L(\hat{\phi}, \hat{\theta}, \sigma^2)) + 2(p+q); \\ BIC &= -2log(L(\hat{\phi}, \hat{\theta}, \sigma^2)) + log(n^*)(p+q) \end{aligned}$$

- The common sample size n^* must be large enough so we can fit an ARMA(p^*,q^*) where p^* and q^* are upper bounds for p and q respectively.
- An ARMA(3,0) and an ARMA(2,0) models were also considered in this paper by Trenberth and Hoar.

```
arma31=arima(x,order=c(3,0,1),n.cond=4)
arma30=arima(x,order=c(3,0,0),n.cond=4)
arma20=arima(x,order=c(2,0,0),n.cond=4)
 arma31$aic
 814.3246
 arma30$aic
 817.3109
 arma20$aic
 823.6855
```

- The minimum value of AIC is for the ARMA(3,1) model.