Time Series Analysis
M.Stat Final Year Students

Lecture 11: ARMA Modelling

Indian Statistical Institute, Kolkata

September 12, 2022

Overview

1. The Dataset
2. Classical Additive Decomposition Model
3. Fitting Some ARMA Models
4. Model Selection Criteria
5. Automatic Selection of the "best" Model
6. Residual Diagnostics for the best Model

The Dataset

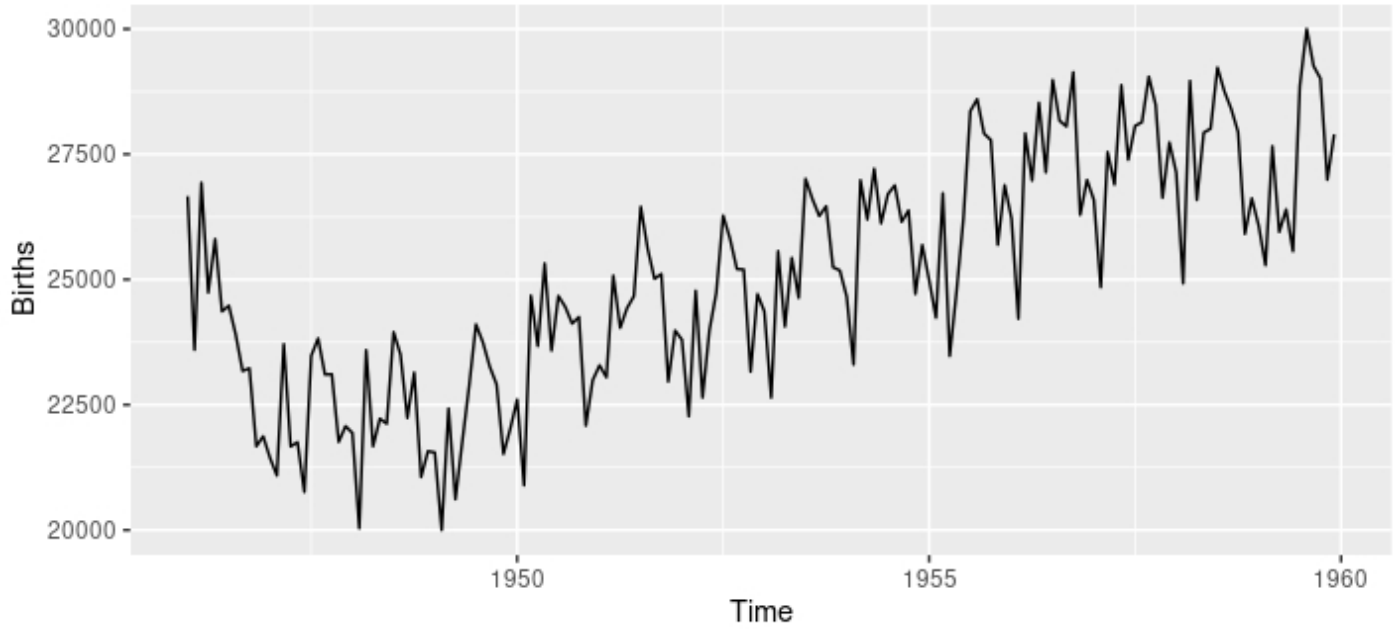
Dataset of number of births per month in New York, from January 1946 to December 1959.

```
> Births <- scan("http://robjhyndman.com/tsdldata/data/nybirths.dat")
Read 168 items
> Births <- ts(Births*1000, frequency = 12, start = c(1946, 1))
> head(Births)
       Jan   Feb   Mar   Apr   May   Jun
1946 26663 23598 26931 24740 25806 24364
```

26663	24477	21439	23479	21937	23950
23598	23901	21089	23824	20035	23504
26931	23175	23709	23105	23590	22238
24740	23227	21669	23110	21672	23142
25806	21672	21752	21759	22222	21059
24364	21870	20761	22073	22123	21573

Plotting the Data

```
autoplot(Births)
```



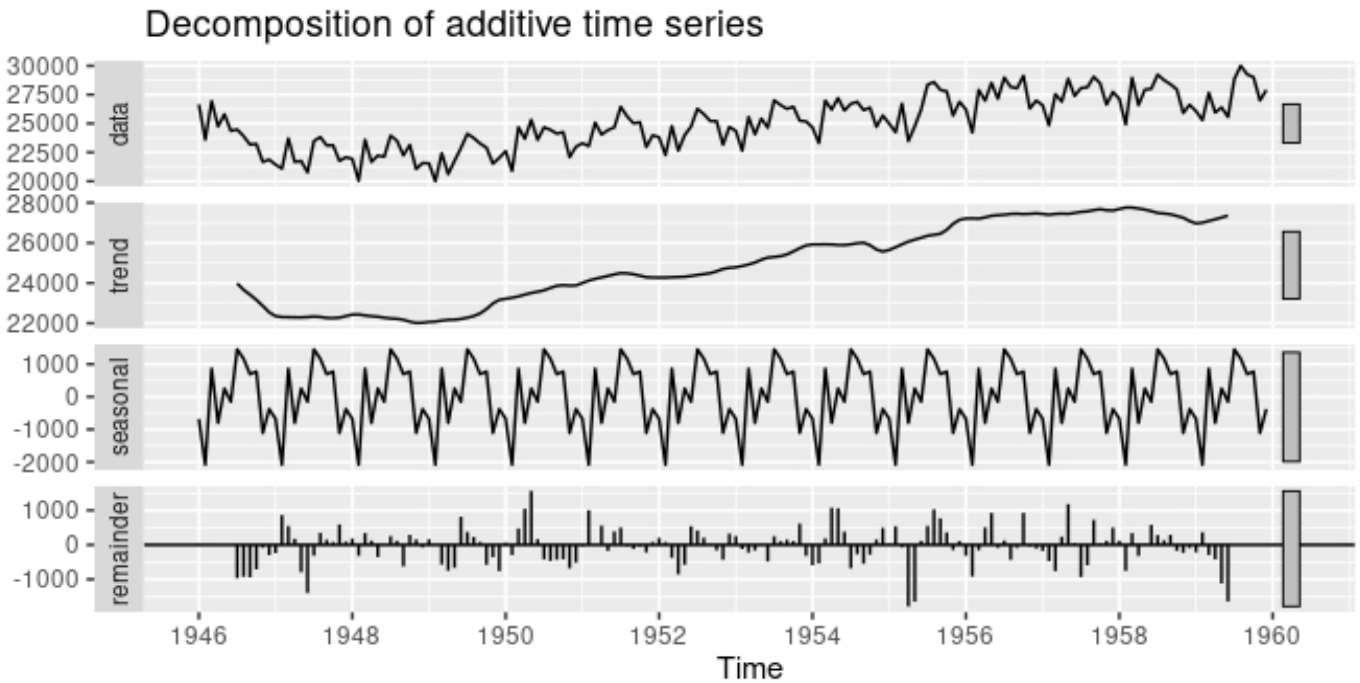
Some seasonal variation in the number of births per month: a peak every mid-year (July-August), and a dip every December-February.

This could be described using an additive model, as the seasonal fluctuations are roughly constant in size over time and do not seem to depend on the level of the time series, and the random fluctuations seem constant over time.

Classical Additive Decomposition Model

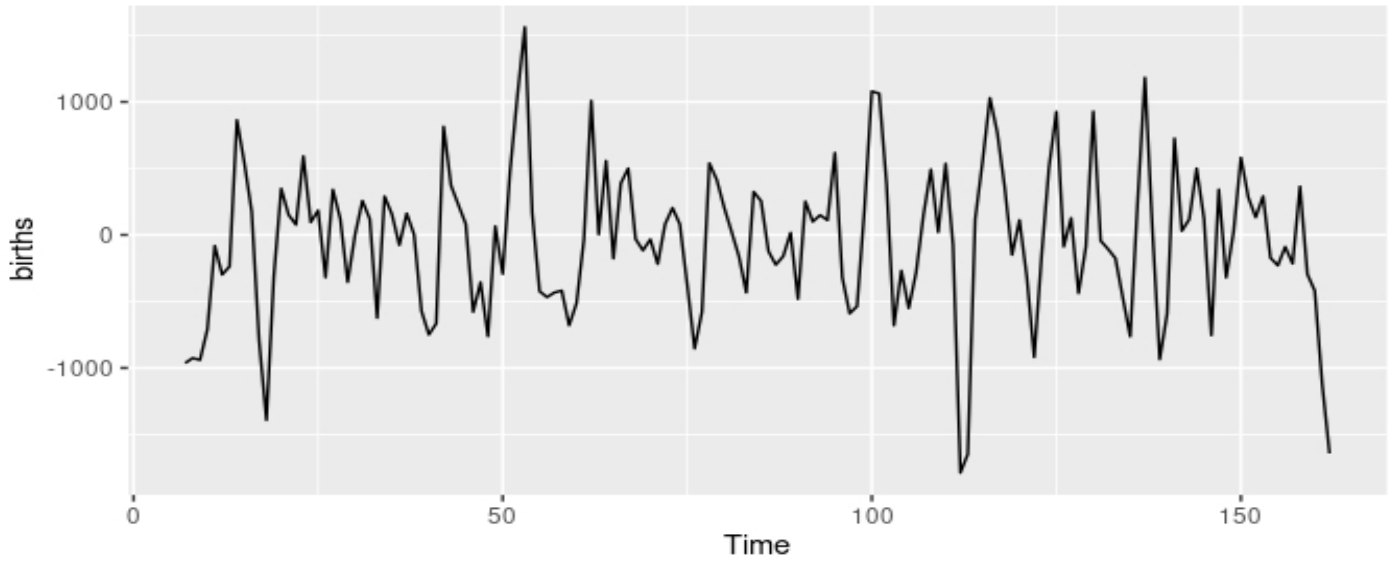
```
birthsComp <- decompose(Births) # classical decomposition
birthsComp
autoplot(birthsComp)
```

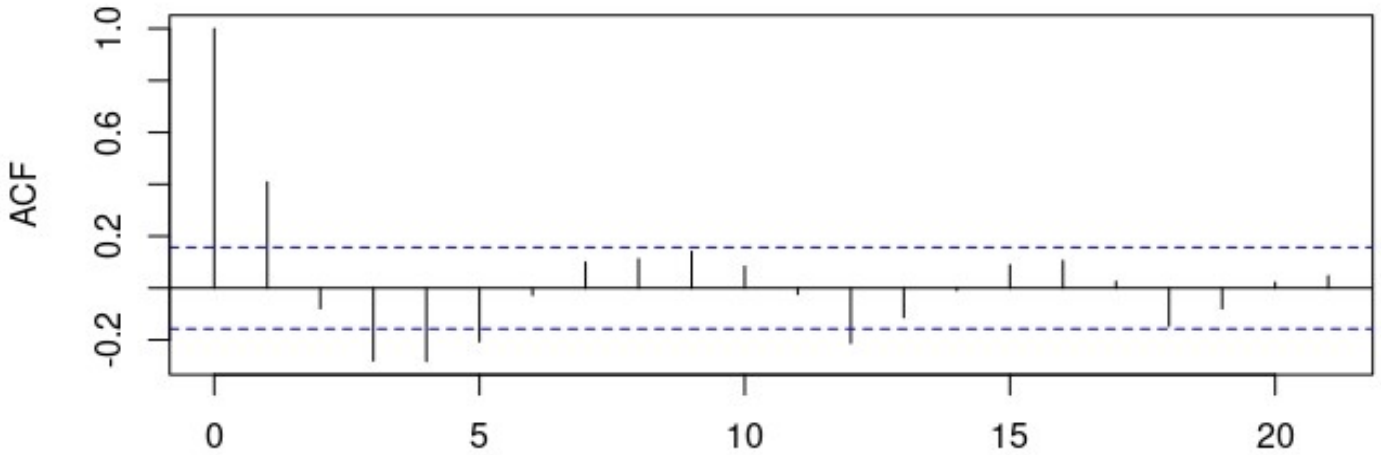
```
    Jan         Feb         Mar         Apr         May         Jun
-677.1947 -2082.9607   862.5232  -801.6787   251.6514  -153.2556
    July        Aug         Sep         Oct         Nov         Dec
1456.0457 1164.5938   691.6162   775.2444 -1109.7652  -376.8197
```



The (Random) Error Component

```
births <- na.omit(ts(birthsComp$random,frequency=1))
autoplot(births)
```



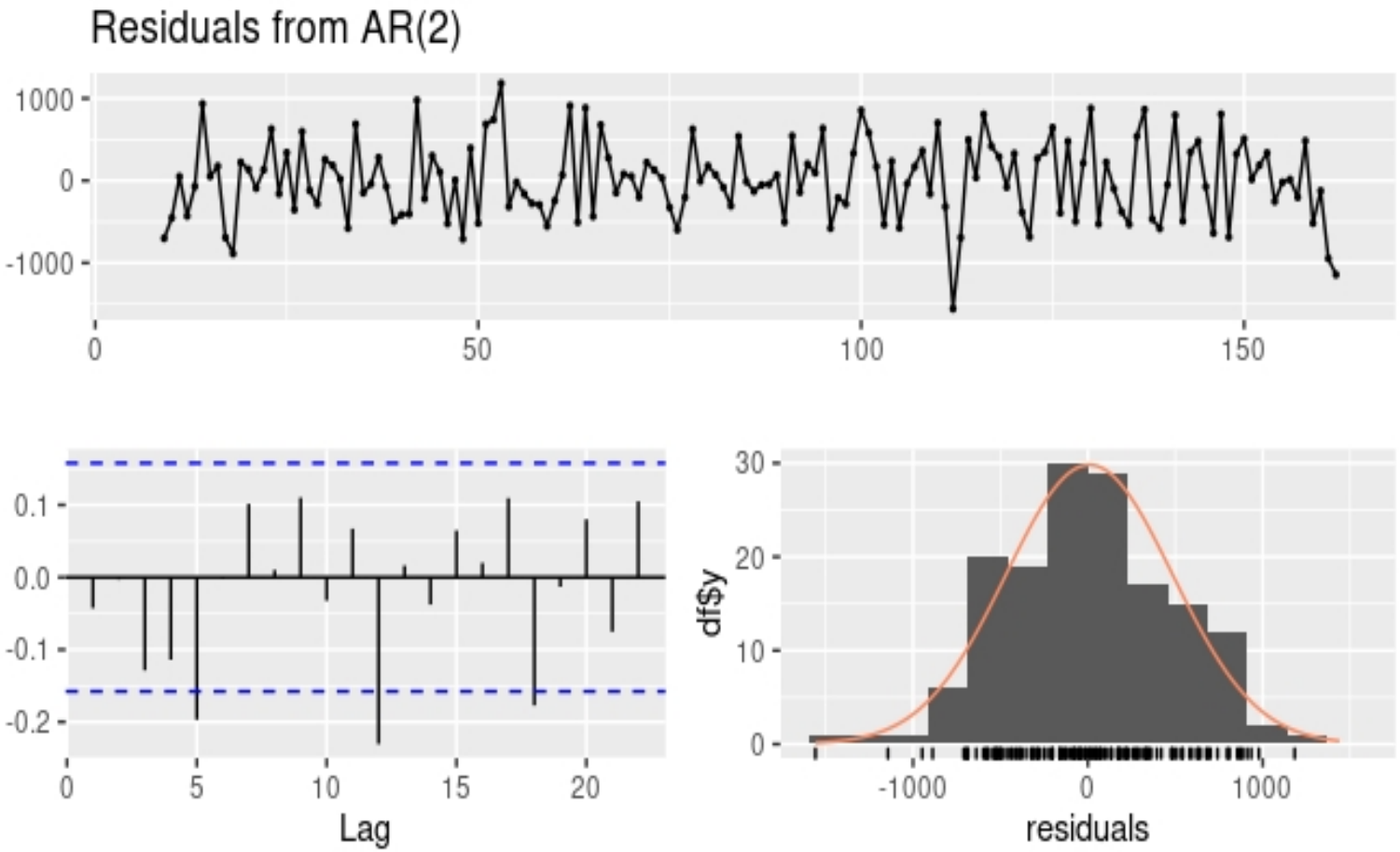


Fitting an AR(2) Model to the Error Component

```
> # fit an AR of order 2
> births_ar2 = ar(births,order.max = 2) #fitting an AR(2) model.
> births_ar2

Call:
ar(x = births, order.max = 2)

Coefficients:
      1         2
 0.5280   -0.2939

Order selected 2  sigma^2 estimated as  237230
> births_ar2$ar # the coefficients of AR(2) model.
[1]  0.5279906 -0.2939002
```

Residual Diagnostics for AR(2)

```
> checkresiduals(births_ar2)
Warning message:
In modeldf.default(object) :
  Could not find appropriate degrees of freedom for this model.
```

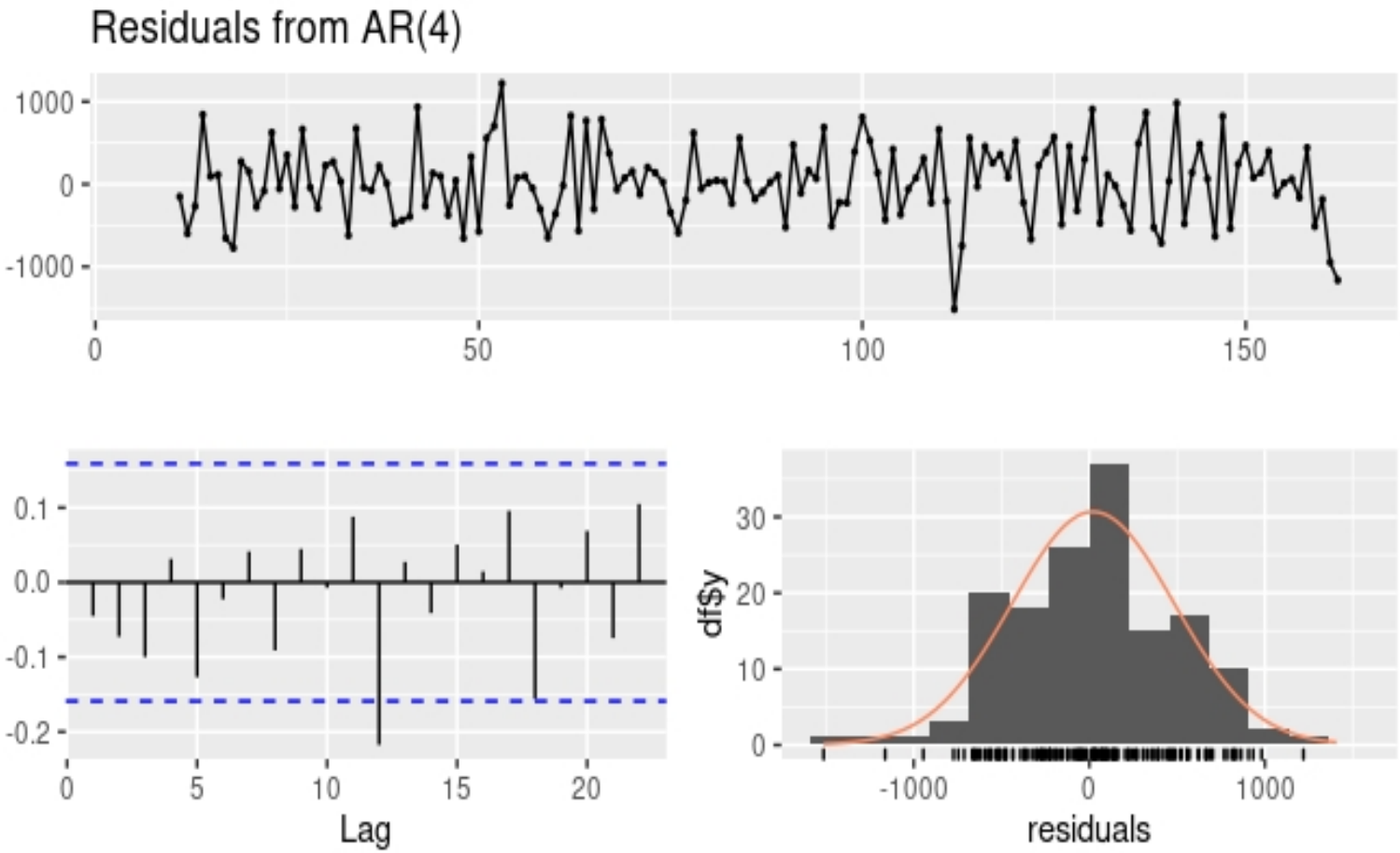


Fitting an AR(4) Model to the Error Component

```
> # fit an AR of order 4
> births_ar4 = ar(births,order.max = 4)
> births_ar4

Call:
ar(x = births, order.max = 4)

Coefficients:
      1         2         3         4
 0.4610   -0.2400   -0.0911   -0.1349

Order selected 4  sigma^2 estimated as  230243
> births_ar4$ar
[1]  0.46103964 -0.24000021 -0.09107613 -0.13492764
```

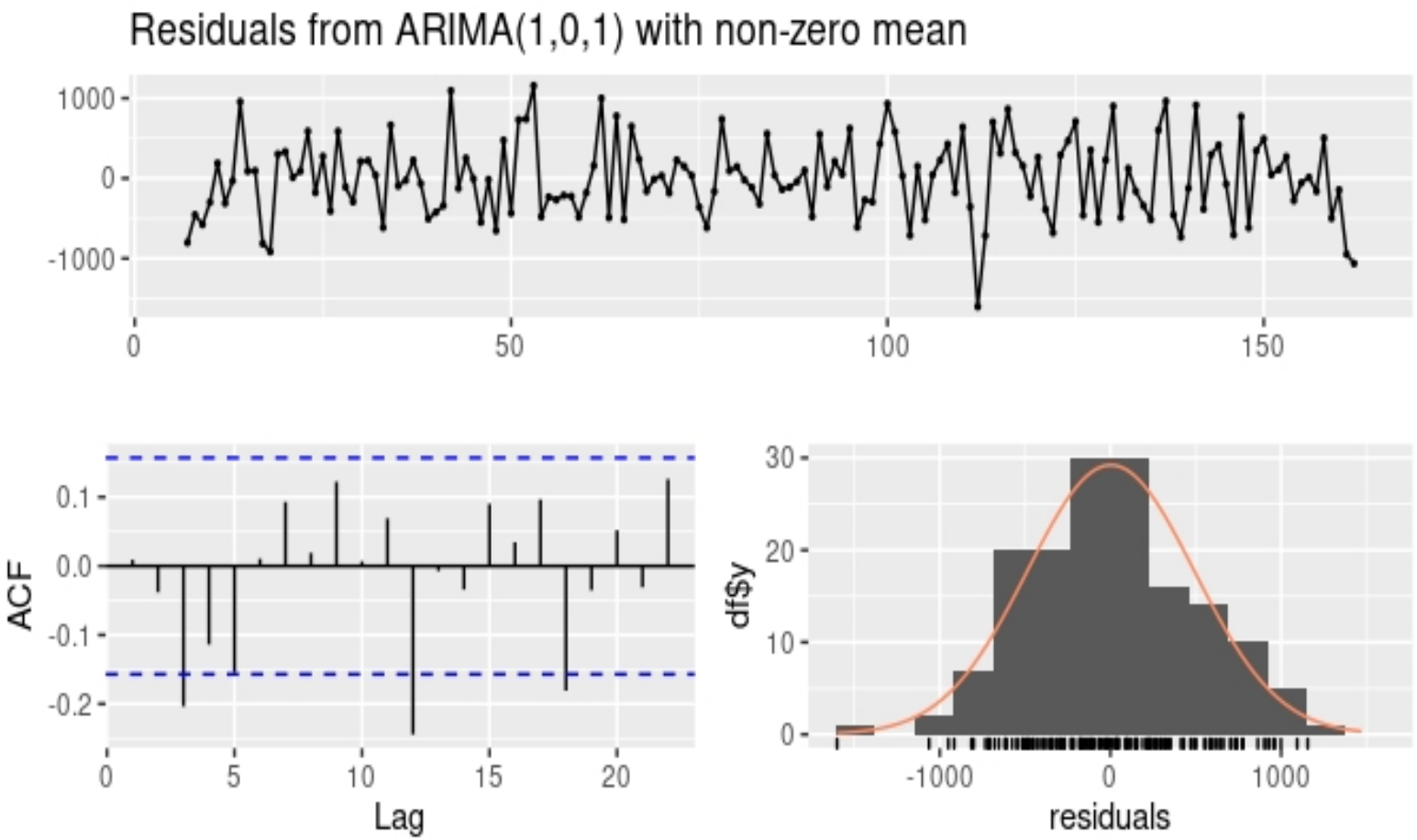
Residual Diagnostics for AR(4)

```
> checkresiduals(births_ar4)
Warning message:
In modeldf.default(object) :
  Could not find appropriate degrees of freedom for this model.
```



Fitting an ARMA(1,1) Model to the Error Component

```
births_arma11 = arima(births, order=c(1,0,1))
checkresiduals(births_arma11)
```

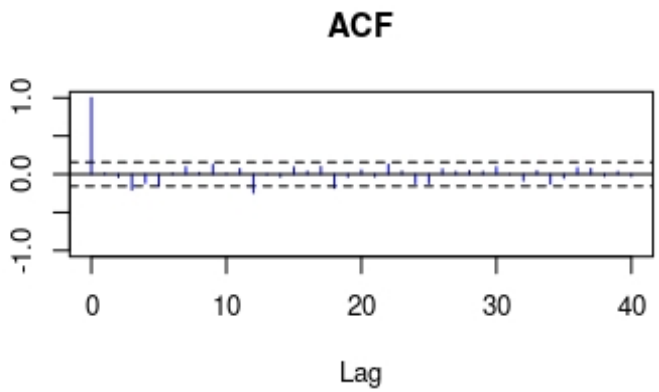


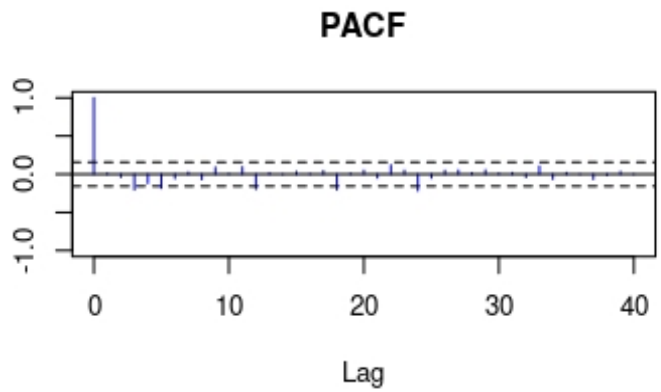
A Bunch of Stationary Tests

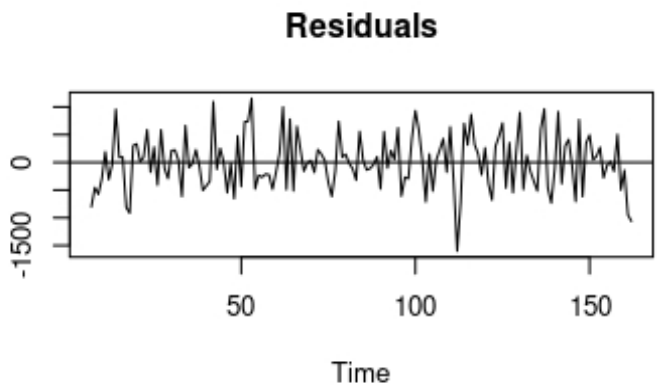
```
> # install.packages("itsmr")
> library(itsmr) # this package contains the test function.
> test(resid(births_arma11))
Null hypothesis: Residuals are iid noise.
Test                   Distribution Statistic   p-value
Ljung-Box Q              Q ~ chisq(20)     38.03    0.0088 *
McLeod-Li Q              Q ~ chisq(20)     21.25    0.3828
Turning points T  (T-102.7)/5.2 ~ N(0,1)       101    0.7502
Diff signs S       (S-77.5)/3.6 ~ N(0,1)        80    0.4895
Rank P           (P-6045)/326.3 ~ N(0,1)      6172    0.6971
```

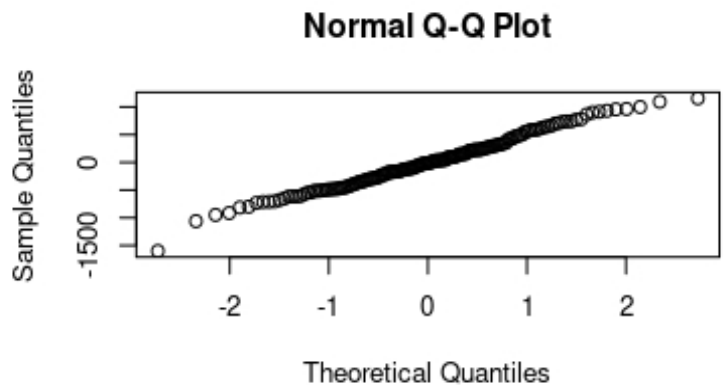
ACF, PACF, Residuals and Q-Q Plot

The **test** function also plots ACF, PACF, residuals, and QQ. The plots can be used to check for stationarity and the other tests check for white noise.









Model Selection Criteria

- $AIC = 2k - 2\ln\left(\hat{L}\right)$.
- $AICc = AIC + \frac{2k(k+1)}{n-k-1}$.
- $BIC = k\ln(n) - 2\ln\left(\hat{L}\right)$.

Automatic Selection of the "best" Model

```
births_best_AIC = auto.arima(births, #the time series data
                  stepwise=FALSE, #explore all models
                  seasonal=FALSE, #take into account seasonality
                  ic="aic", #information criterion
                  trace=TRUE #show fitting models)
```

```
ARIMA(0,0,0) with zero mean     : 2415.912
ARIMA(0,0,0) with non-zero mean : 2416.964
ARIMA(0,0,1) with zero mean     : 2380.696
ARIMA(0,0,1) with non-zero mean : 2381.945
ARIMA(0,0,2) with zero mean     : 2380.094
ARIMA(0,0,2) with non-zero mean : 2381.406

...

ARIMA(4,0,1) with non-zero mean : Inf
ARIMA(5,0,0) with zero mean     : 2363.77
ARIMA(5,0,0) with non-zero mean : 2365.262

Now re-fitting the best model(s) without approximations...



Best model: ARIMA(2,0,1) with zero mean
```

The "best" Model

```
> births_best_AIC
Series: births
ARIMA(2,0,1) with zero mean

Coefficients:
          ar1      ar2      ma1
      1.2762  -0.6069  -0.9124
s.e.  0.0851   0.0721   0.0803

sigma^2 = 212964:  log likelihood = -1177.59
AIC=2363.17   AICc=2363.44   BIC=2375.37
```

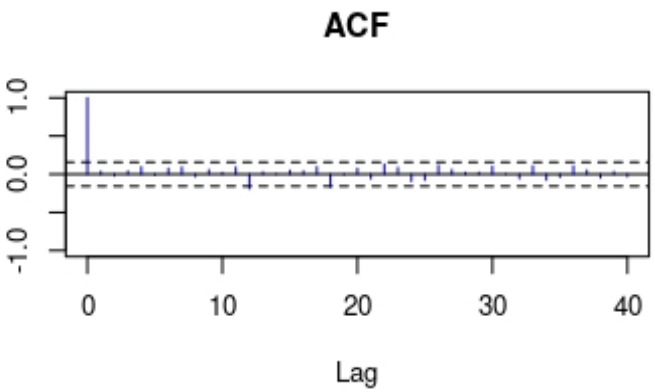
The "best" Model

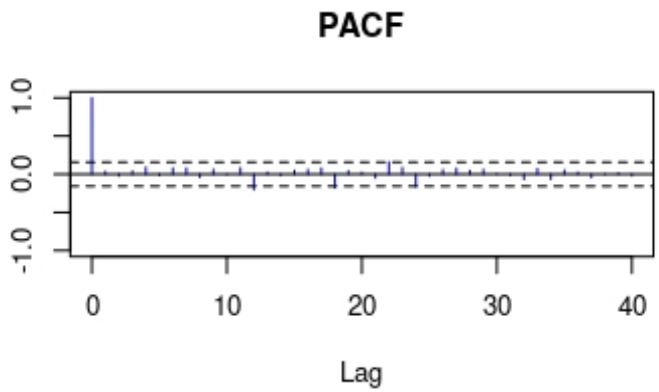
```
> births_best_AICc
Series: births
ARIMA(2,0,1) with zero mean

Coefficients:
         ar1      ar2      ma1
      1.2762  -0.6069  -0.9124
s.e.  0.0851   0.0721   0.0803

sigma^2 = 212964:  log likelihood = -1177.59
AIC=2363.17   AICc=2363.44   BIC=2375.37
> births_best_BIC
Series: births
ARIMA(2,0,1) with zero mean

Coefficients:
         ar1      ar2      ma1
      1.2762  -0.6069  -0.9124
s.e.  0.0851   0.0721   0.0803

sigma^2 = 212964:  log likelihood = -1177.59
AIC=2363.17   AICc=2363.44   BIC=2375.37
```

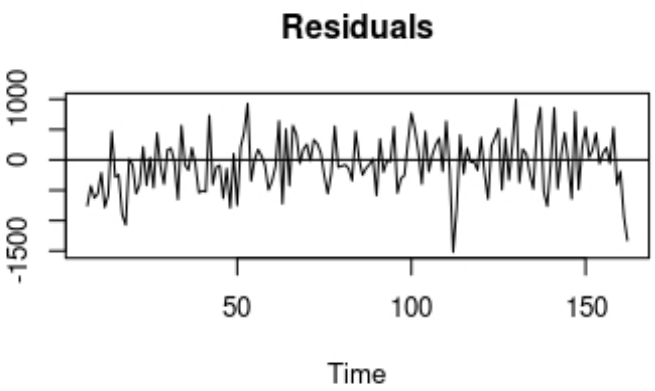
Residual Diagnostics for the best Model

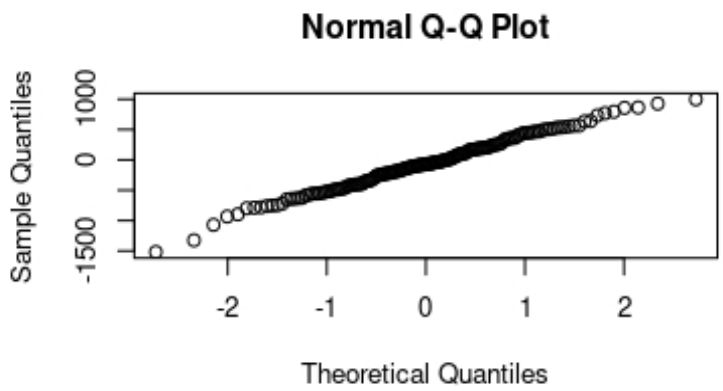
```
> test(resid(births_best_AIC))
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                   Q ~ chisq(20)     20.87    0.4051
McLeod-Li Q                   Q ~ chisq(20)      9.86    0.9706
Turning points T  (T-102.7)/5.2 ~ N(0,1)       101    0.7502
Diff signs S       (S-77.5)/3.6 ~ N(0,1)        81    0.3332
Rank P           (P-6045)/326.3 ~ N(0,1)      6899    0.0089 *
```

ACF, PACF, Residuals and Q-Q Plot









Thank You!