universität
innsbruck

Time Series Analysis

Model Diagnostics

Model Diagnostics

Contents

Contents

Model Diagnostics

Residual Analysis

Residual analysis

Idea: Residuals $\widehat{e}_t$ from correctly specified model should be similar to white noise, i.e., independent, zero mean, and constant variance (and normal).

Residuals:

- By-product of model estimation (see Chapter 7).
- Can also be seen as one-step-ahead forecast errors (see Chapter 9).

Example: AR(2) with constant term.

$$
\begin{aligned}
Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_0 + e_t \\
\widehat{Y}_t &= \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2} + \hat{\theta}_0 \\
\widehat{e}_t &= Y_t - \widehat{Y}_t
\end{aligned}
$$

Residual analysis: Time series plot

Illustration: MA(1) model for oil price returns.

```
R> data("oil.price", package = "TSA")
R> oil <- diff(log(oil.price))
R> oil_ma1 <- arima(oil, order = c(0, 0, 1), include.mean = FALSE)
R> oil_ma1
```

```
Call:
arima(x = oil, order = c(0, 0, 1), include.mean = FALSE)

Coefficients:
        ma1
      0.296
s.e.  0.069

sigma^2 estimated as 0.00669:  log likelihood = 260.3,  aic = -516.6
```

Residual analysis: Time series plot

Extract and plot residuals $\widehat{e}_t$:

```
R> plot(residuals(oil_ma1))
R> abline(h = 0, col = "slategray")
```

Extract and plot standardized residuals $\widehat{e}_t/\hat{\sigma}$:

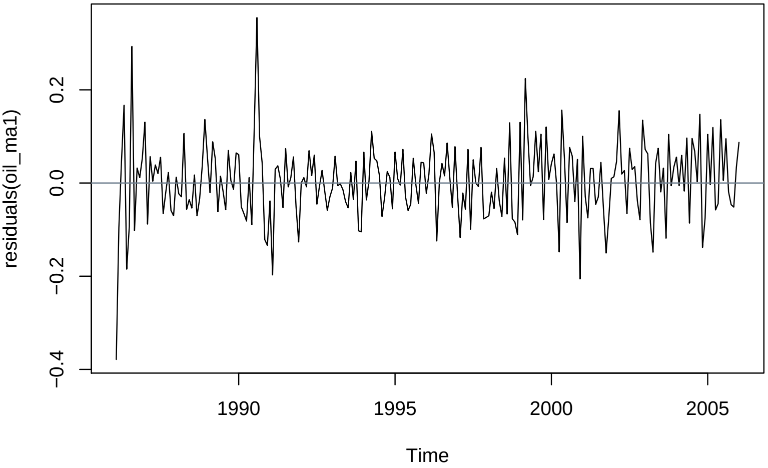
```
R> oil_res <- residuals(oil_ma1) / sqrt(oil_ma1$sigma2)
R> plot(oil_res)
R> abline(h = 0, col = "slategray")
```

Plot standardized residuals in "histogram" style:

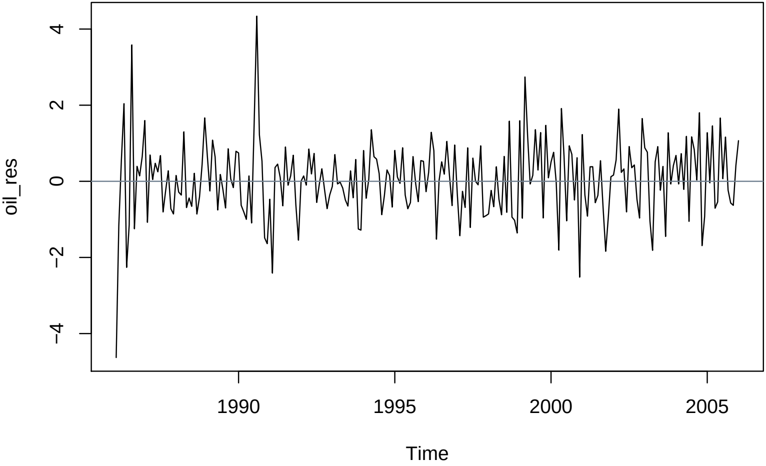
```
R> plot(oil_res, type = "h")
R> abline(h = 0, col = "slategray")
```

Interpretation: Mostly ok except for two outliers with absolute values in excess of 4 in Feb 1986 and Aug 1990.

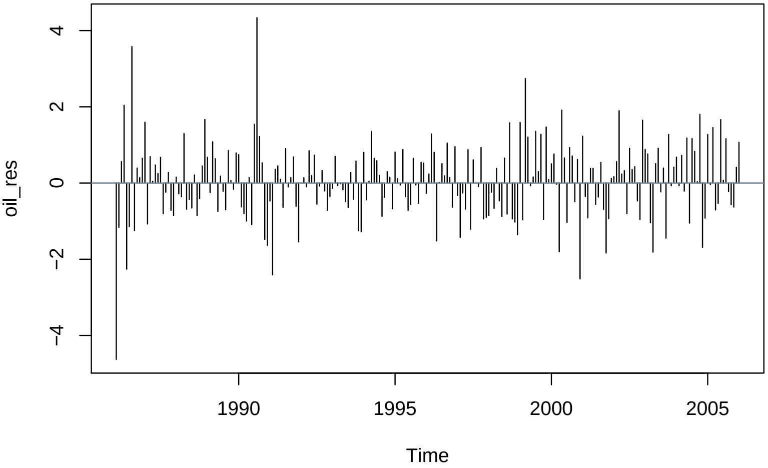
Residual analysis: Time series plot



Residual analysis: Time series plot



Residual analysis: Time series plot



Residual analysis: Distribution

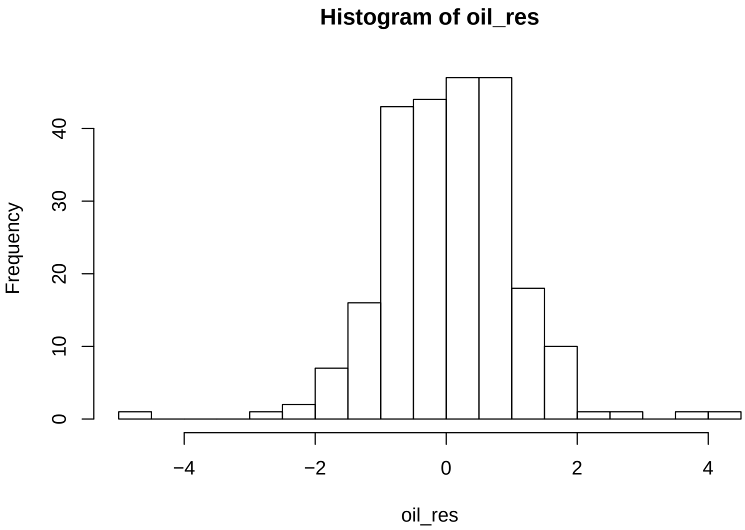
Outliers can also be seen in histogram (collapsed over time):

```
R> oil_res <- residuals(oil_ma1) / sqrt(oil_ma1$sigma2)
R> hist(oil_res, breaks = 20)
```

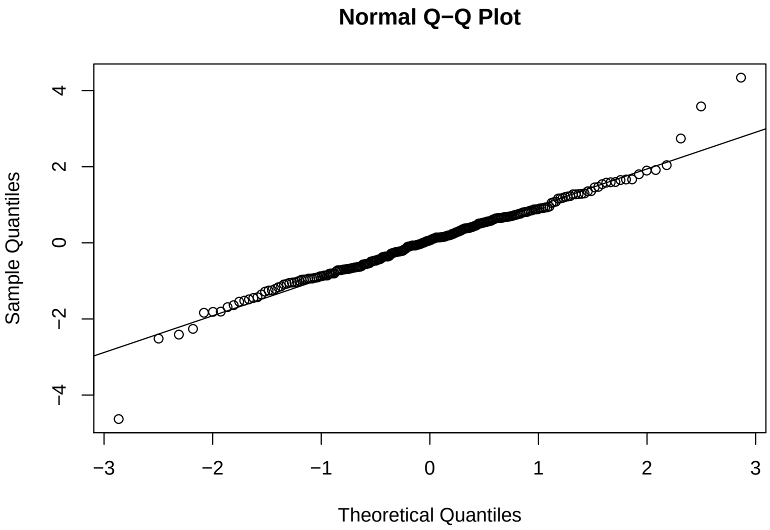
Or in QQ plot for checking (approximate) normality:

```
R> qqnorm(oil_res)
R> qqline(oil_res)
```

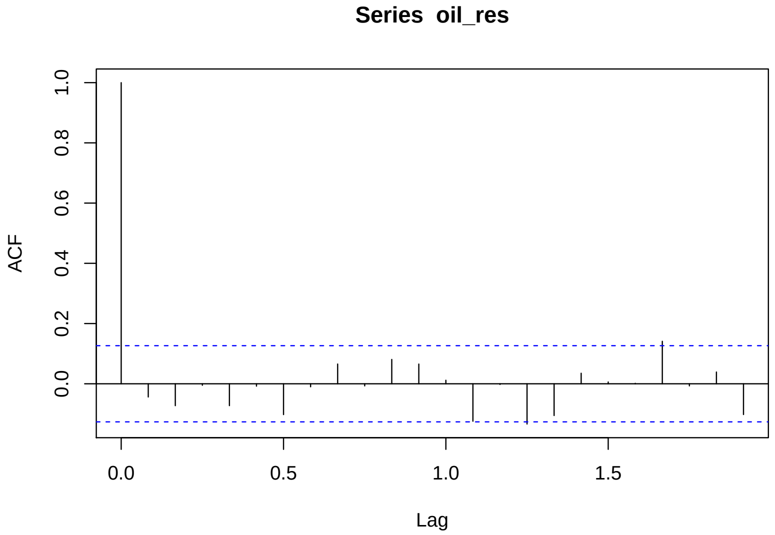
Residual analysis: Histogram



Residual analysis: QQ plot



Residual analysis: Autocorrelation



Residual analysis: Autocorrelation

Crucial question: Does the model capture the autocorrelation of the data appropriately? Is the associated noise term independent?

Idea: Consider empirical autocorrelation function of residuals $\widehat{e}_t$, denoted $\widehat{r}_k$.

Recall: For large n, empirical autocorrelations of white noise process are approximately uncorrelated with zero mean and variance $1/n$.

Problem: Residuals $\widehat{e}_t$ from a correctly specified and efficiently estimated ARIMA model have somewhat different properties.

Residual analysis: Autocorrelation

Known: Durbin-Watson test for linear regression models *without* lagged dependent variable. Cannot be applied to residuals from ARMA model.

Solution: Derive (asymptotic) distribution of ARMA residuals, due to Box and Pierce (1970).

Properties:

- Residuals $\widehat{e}_t$ are approximately normal with zero means.
- For small k, variance of $\widehat{r}_k$ can be substantially less than $1/n$.
- For small k and j, estimates $\widehat{r}_k$ and $\widehat{r}_j$ can be highly correlated.
- For larger lags, correlations are approximately uncorrelated and have variance $1/n$.

Residual analysis: Autocorrelation

Example: Correctly specified, efficiently estimated AR(1). For large n

$$\begin{aligned}
\mathrm{Var}(\widehat{r}_1) &\approx \frac{\phi^2}{n} \\
\mathrm{Var}(\widehat{r}_k) &\approx \frac{1 \, - \, (1-\phi^2)\,\phi^{2k-2}}{n} \\
\mathrm{Cor}(\widehat{r}_1, \widehat{r}_k) &\approx -\mathrm{sign}(\phi)\,\frac{(1-\phi^2)\,\phi^{k-2}}{1 \, - \, (1-\phi^2)\,\phi^{2k-2}}
\end{aligned}$$

Residual analysis: Autocorrelation

Function for computing $n \cdot \text{Var}(\widehat{r}_k)$ given k and ϕ:

```
R> nvark <- function(k = 1, phi = 0)
+    1 - (1 - phi^2) * phi^(2 * k - 2)
```

Analogously, function for $\text{Cor}(\widehat{r}_1, \widehat{r}_k)$ given k and ϕ:

```
R> cor1k <- function(k = 1, phi = 0) - sign(phi) *
+    ((1 - phi^2) * phi^(k - 2)) / nvark(phi = phi, k = k)
```

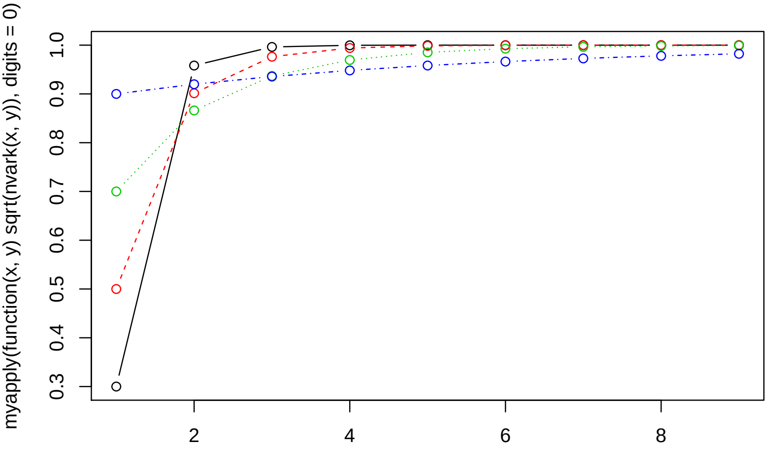
Convenience function for computation across k and ϕ values, including prettified labeling:

```
R> myapply <- function(FUN, lag = 1:9,
+    ar = c(0.3, 0.5, 0.7, 0.9), digits = 2)
+  {
+    rval <- outer(lag, ar, FUN)
+    dimnames(rval) <- list(lag, ar)
+    if(digits > 0) rval <- round(rval, digits = digits)
+    rval
+  }
```

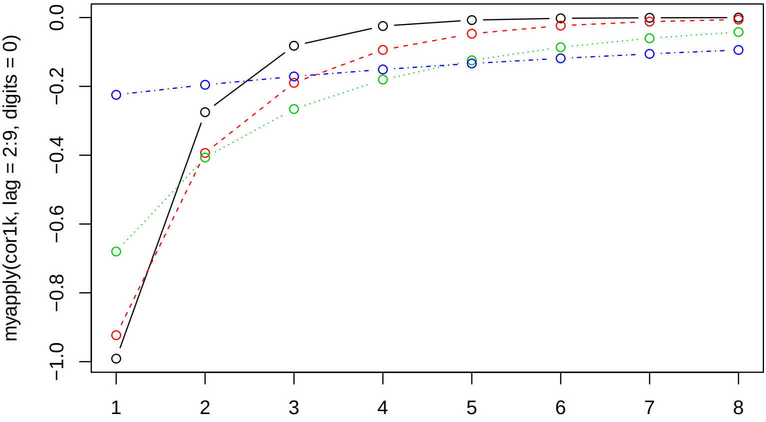
Residual analysis: Autocorrelation

```
R> myapply(function(x, y) sqrt(nvark(x, y)))
   0.3  0.5  0.7  0.9
1 0.30 0.50 0.70 0.90
2 0.96 0.90 0.87 0.92
3 1.00 0.98 0.94 0.94
4 1.00 0.99 0.97 0.95
5 1.00 1.00 0.99 0.96
6 1.00 1.00 0.99 0.97
7 1.00 1.00 1.00 0.97
8 1.00 1.00 1.00 0.98
9 1.00 1.00 1.00 0.98
R> myapply(cor1k, lag = 2:9)
    0.3   0.5   0.7   0.9
2 -0.99 -0.92 -0.68 -0.22
3 -0.28 -0.39 -0.41 -0.20
4 -0.08 -0.19 -0.27 -0.17
5 -0.02 -0.09 -0.18 -0.15
6 -0.01 -0.05 -0.12 -0.13
7  0.00 -0.02 -0.09 -0.12
8  0.00 -0.01 -0.06 -0.11
9  0.00 -0.01 -0.04 -0.09
```

Residual analysis: Autocorrelation



Residual analysis: Autocorrelation



Residual analysis: Autocorrelation

Similarly: Correctly specified, efficiently estimated AR(2). For large n, if the parameters are not too close to the stationarity boundary:

$$\begin{aligned}
\mathrm{Var}(\widehat{r}_1) &\approx \frac{\phi_2^2}{n} \\
\mathrm{Var}(\widehat{r}_2) &\approx \frac{\phi_2^2 + \phi_1^2(1+\phi_2)^2}{n} \\
\mathrm{Var}(\widehat{r}_k) &\approx \frac{1}{n} \qquad (k \geq 3)
\end{aligned}$$

Analogously: MA(1) and MA(2) case with θs instead of ϕs.

Moreover: Results for general ARMA models in Box and Pierce (1970) and McLeod (1978).

Residual analysis: Ljung-Box test

Idea: Test formally for white noise by aggregating autocorrelation across lags.

Test: Box and Pierce (1970) show that for a correct ARMA(p, q) specification

$$Q \; = \; n\,(\widehat{r}_1^2 \; + \; \widehat{r}_2^2 \; + \; \ldots \; + \; \widehat{r}_K^2)$$

is asymptotically $\chi^2_{K-(p+q)}$ distributed.

Under null hypothesis of white noise Q should be close to zero. Under alternative, Q is inflated.

Maximal lag K is typically chosen ad hoc so that AR(∞) weights ψ_j are sufficiently small for $j > K$. Alternatively consider range $k = 1, \ldots, K$.

Residual analysis: Ljung-Box test

Problem: Ljung and Box (1978) show that the asymptotic distribution approximates the finite sample distribution only poorly, even for n around 100.

Solution: Employ finite sample correction

$$Q^* = n\,(n+2)\,\left(\frac{\widehat{r}_1^2}{n-1} + \frac{\widehat{r}_2^2}{n-2} + \ldots + \frac{\widehat{r}_K^2}{n-K}\right)$$

As $n+2 > n-k$ holds for all k, $Q^* > Q$ and thus has more power against model deviations.

Name: Ljung-Box test, also known as modified Box-Pierce test.

In R: `Box.test()`.

Residual analysis: Ljung-Box test

Illustration: MA(1) model for oil price returns (i.e., $p + q = 1$).

```
R> Box.test(oil_res, lag = 5, fitdf = 1)

        Box-Pierce test

data:  oil_res
X-squared = 3, df = 4, p-value = 0.6

R> Box.test(oil_res, lag = 5, fitdf = 1, type = "Ljung-Box")

        Box-Ljung test

data:  oil_res
X-squared = 3.1, df = 4, p-value = 0.5

R> n <- length(oil)
R> n * sum(acf(oil_res, plot = FALSE)$acf[2:6]^2)
[1] 3.023
R> n * (n + 2) * sum(acf(oil_res, plot = FALSE)$acf[2:6]^2/(n - 1:5))
[1] 3.083
```

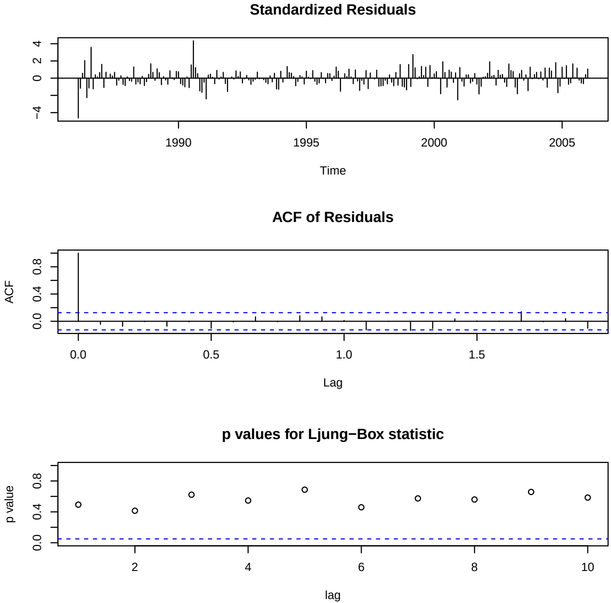
Residual analysis: Summary

In practice:

- Assess time series of residuals $\widehat{e}_t$ for patterns in mean or variance or outliers.
- Inspect empirical autocorrelation $\widehat{r}_k$ of residuals for remaining autocorrelations.
- Test null hypothesis of white noise using aggregate autocorrelations up to some lag K.

In R: `tsdiag()` for “`Arima`” objects fitted by `arima()`

Residual analysis: Summary



Model Diagnostics

Overfitting and Parameter Redundancy

Overfitting and parameter redundancy

Idea: Assess whether a somewhat more general ARMA(p^*, q^*) improves on a given ARMA(p, q) model with $p^* > p$ and/or $q^* > q$.

Test: Likelihood ratio (LR) test compares associated fitted log-likelihoods $\log L(\hat{\beta}^*)$ and $\log L(\hat{\beta})$. The LR test statistic

$$2 \cdot \left\{ \log L(\hat{\beta}^*) \; - \; \log L(\hat{\beta}) \right\}$$

is asymptotically $\chi^2_{p^*+q^*-(p+q)}$ distributed.

Alternatively: Wald test compares the associated parameter estimates $\hat{\beta}^*$ and $\hat{\beta}$.

Intuition: Overfitted parameter estimates $\hat{\beta}^*$ should be close to zero for additional parameters and close to $\hat{\beta}$ otherwise. However, intuition is not always correct (see below).

Overfitting and parameter redundancy

In R: `lrtest()` and `coeftest()` from *lmtest*.

Remark: For pretty printing in `lrtest()` employ

```
R> myname <- function(x) x$call
```

Illustration: Alternative ARMA(p, q) specification with and without intercepts for oil price returns.

```
R> oil_arma1 <- arima(oil, order = c(0, 0, 2), include.mean = FALSE)
R> oil_arma2 <- arima(oil, order = c(1, 0, 1), include.mean = FALSE)
R> oil_arma3 <- arima(oil, order = c(0, 0, 1), include.mean = TRUE)
R> oil_arma4 <- arima(oil, order = c(1, 0, 2), include.mean = TRUE)
```

Overfitting and parameter redundancy

```
R> oil_arma1

Call:
arima(x = oil, order = c(0, 0, 2), include.mean = FALSE)

Coefficients:
        ma1      ma2
      0.269  -0.094
s.e.  0.068   0.075

sigma^2 estimated as 0.00664:  log likelihood = 261.1,  aic = -516.2
R> coeftest(oil_arma1)

z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ma1   0.2690     0.0680    3.95  7.7e-05
ma2  -0.0941     0.0746   -1.26     0.21
```

Overfitting and parameter redundancy

```
R> lrtest(oil_ma1, oil_arma1, name = myname)
Likelihood ratio test

Model 1: arima(x = oil, order = c(0, 0, 1), include.mean = FALSE)
Model 2: arima(x = oil, order = c(0, 0, 2), include.mean = FALSE)
  #Df LogLik Df Chisq Pr(>Chisq)
1   2    260
2   3    261  1  1.58       0.21
```

Overfitting and parameter redundancy

```
R> oil_arma2
Call:
arima(x = oil, order = c(1, 0, 1), include.mean = FALSE)

Coefficients:
          ar1    ma1
      -0.299  0.570
s.e.   0.201  0.172

sigma^2 estimated as 0.00664:  log likelihood = 261.1,  aic = -516.2
R> coeftest(oil_arma2)
z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ar1   -0.299      0.201   -1.49  0.13696
ma1    0.570      0.172    3.31  0.00094
```

Overfitting and parameter redundancy

```
R> lrtest(oil_ma1, oil_arma2, name = myname)
Likelihood ratio test

Model 1: arima(x = oil, order = c(0, 0, 1), include.mean = FALSE)
Model 2: arima(x = oil, order = c(1, 0, 1), include.mean = FALSE)
  #Df LogLik Df Chisq Pr(>Chisq)
1   2    260
2   3    261  1  1.64        0.2
```

Overfitting and parameter redundancy

```
R> oil_arma3
Call:
arima(x = oil, order = c(0, 0, 1), include.mean = TRUE)

Coefficients:
        ma1  intercept
      0.294      0.004
s.e.  0.070      0.007

sigma^2 estimated as 0.00668:  log likelihood = 260.5,  aic = -514.9
R> coeftest(oil_arma3)
z test of coefficients:

          Estimate Std. Error z value Pr(>|z|)
ma1        0.29390    0.06955    4.23  2.4e-05
intercept  0.00406    0.00682    0.59     0.55
```

Overfitting and parameter redundancy

```
R> lrtest(oil_ma1, oil_arma3, name = myname)
Likelihood ratio test

Model 1: arima(x = oil, order = c(0, 0, 1), include.mean = FALSE)
Model 2: arima(x = oil, order = c(0, 0, 1), include.mean = TRUE)
  #Df LogLik Df Chisq Pr(>Chisq)
1   2    260
2   3    260  1  0.35       0.55
```

Overfitting and parameter redundancy

```
R> oil_arma4
Call:
arima(x = oil, order = c(1, 0, 2), include.mean = TRUE)

Coefficients:
        ar1      ma1      ma2  intercept
      0.868  -0.607  -0.320      0.004
s.e.  0.083   0.100   0.068      0.003

sigma^2 estimated as 0.00655:  log likelihood = 262.7,  aic = -515.4
R> coeftest(oil_arma4)
z test of coefficients:

          Estimate Std. Error z value Pr(>|z|)
ar1        0.86801    0.08289   10.47  < 2e-16
ma1       -0.60671    0.10007   -6.06  1.3e-09
ma2       -0.31970    0.06756   -4.73  2.2e-06
intercept  0.00447    0.00309    1.45     0.15
```

Overfitting and parameter redundancy

```
R> lrtest(oil_ma1, oil_arma4, name = myname)
Likelihood ratio test

Model 1: arima(x = oil, order = c(0, 0, 1), include.mean = FALSE)
Model 2: arima(x = oil, order = c(1, 0, 2), include.mean = TRUE)
  #Df LogLik Df Chisq Pr(>Chisq)
1   2    260
2   5    263  3  4.85       0.18
```

Overfitting and parameter redundancy

Problem: For overfitted ARMA(p^*, q^*) models, parameters may be redundant, i.e., we have *lack of identifiability*.

Illustration: ARMA(1, 2) model.

$$Y_t \;=\; \phi Y_{t-1} \;+\; e_t \;-\; \theta_1 e_{t-1} \;-\; \theta_2 e_{t-2}$$

Thus, at time $t-1$ instead of t

$$Y_{t-1} \;=\; \phi Y_{t-2} \;+\; e_{t-1} \;-\; \theta_1 e_{t-2} \;-\; \theta_2 e_{t-3}$$

Multiplying both sides with some constant c and subtracting it from the first equation yields after rearrangement

$$Y_t-(\phi+c)Y_{t-1}+\phi c Y_{t-2} \;=\; e_t-(\theta_1+c)e_{t-1}-(\theta_2-\theta_1 c)e_{t-2}+\theta_2 c e_{t-3}$$

i.e., apparently an ARMA(2, 3) process.

Overfitting and parameter redundancy

Note that the following factorizations hold for the associated characteristic polynomials

$$\begin{aligned} 1-(\phi+c)x+\phi c x^2 &= (1-\phi x)\,(1-cx) \\ 1-(\theta_1+c)x-(\theta_2-\theta_1 c)x^2+\theta_2 c x^3 &= (1-\theta_1 x-\theta_2 x^2)\,(1-cx) \end{aligned}$$

Thus:

- AR and MA characteristic polynomials have common factor $(1-cx)$, i.e., common root $1/c$.
- Parameters in ARMA(2, 3) specification are not unique, c is completely arbitrary.
- Can occur in particular when increasing p and q simultaneously, i.e., potentially including redundant terms in both AR and MA parts.

Overfitting and parameter redundancy

However: Fitted log-likelihood is unique, i.e., information criteria and LR tests are not affected.

More generally: In backshift notation. If

$$\phi(B)Y_t \;=\; \theta(B)e_t$$

is a correct model, then so is

$$(1 - cB)\,\phi(B)Y_t = (1 - cB)\,\theta(B)e_t$$

for any constant c.

To have a unique parametrization in an ARMA model, all common factors in AR and MA characteristic polynomials must be canceled out.