Ch 8. MODEL DIAGNOSTICS

Time Series Analysis

Model diagnostics is concerned with testing the goodness of fit of a model and, if the fit is poor, suggesting appropriate modifications. We shall present two complementary approaches: analysis of residuals from the fitted model and analysis of overparameterized models.

8.1 Residual Analysis

For an AR(p) model: $Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + \theta_0 + e_t$, after obtaining the estimated coefficients $\hat{\phi}_k$ and $\hat{\theta}_0$, the residuals are defined as

$$\hat{e}_t = Y_t - \sum_{k=1}^{p} \hat{\phi}_k Y_{t-k} - \hat{\theta}_0.$$

For an *invertible* ARMA(p,q) model containing MA terms, we use the inverted infinite AR form of the model to define residuals:

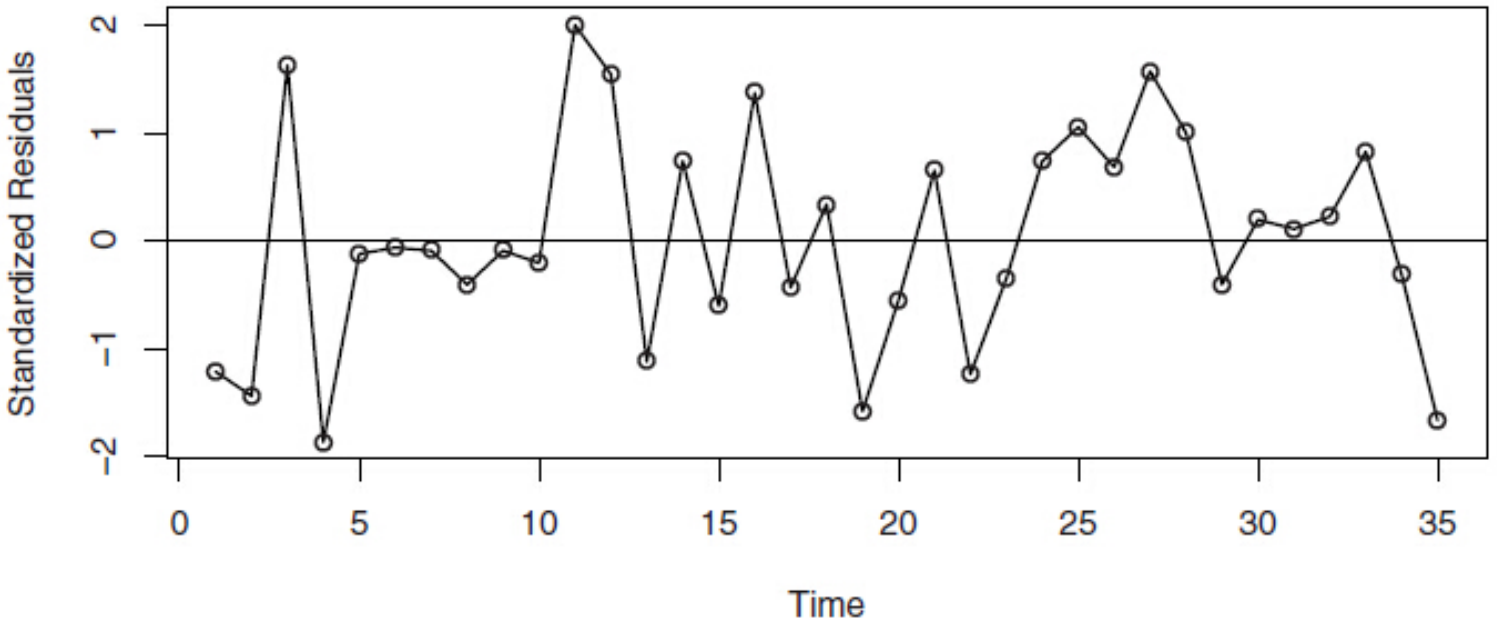
$$\begin{aligned} & Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \cdots + e_t \\ \Rightarrow \quad & \hat{e}_t = \hat{Y}_t - \hat{\pi}_1 Y_{t-1} - \hat{\pi}_2 Y_{t-2} - \hat{\pi}_3 Y_{t-3} - \cdots \\ \text{or} \quad & \text{residual} = \text{actual} - \text{predicted}. \end{aligned}$$

The $\hat{\pi}$'s are calculated implicitly as functions of the estimated ϕ's and θ's.

If the model is correctly specified and the parameter estimates are reasonably close to the true values, then the residuals should have nearly the properties of white noise.

8.1.1. Plots of the Residuals

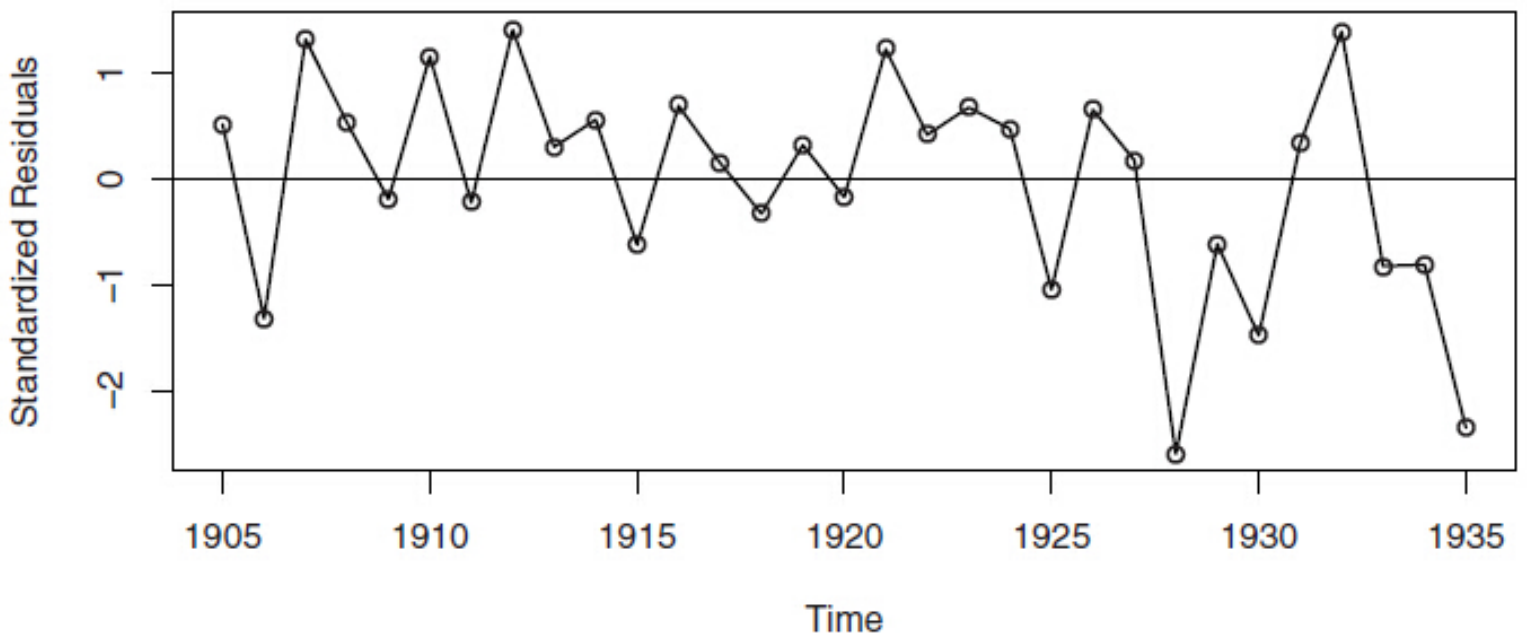
Exhibit 8.1 Standardized Residuals from AR(1) Model of Color




```
> win.graph(width=4.875,height=3,pointsize=8)
> data(color)
> m1.color=arima(color,order=c(1,0,0)); m1.color
> plot(rstandard(m1.color),ylab ='Standardized Residuals',
    type='o'); abline(h=0)
```

The parameters of the AR(1) model were estimated using maximum likelihood. The plot shows a rectangular scatter around a zero horizontal level with no trend, which supports the model.

Exhibit 8.2 Standardized Residuals from AR(3) Model for Sqrt(Hare)

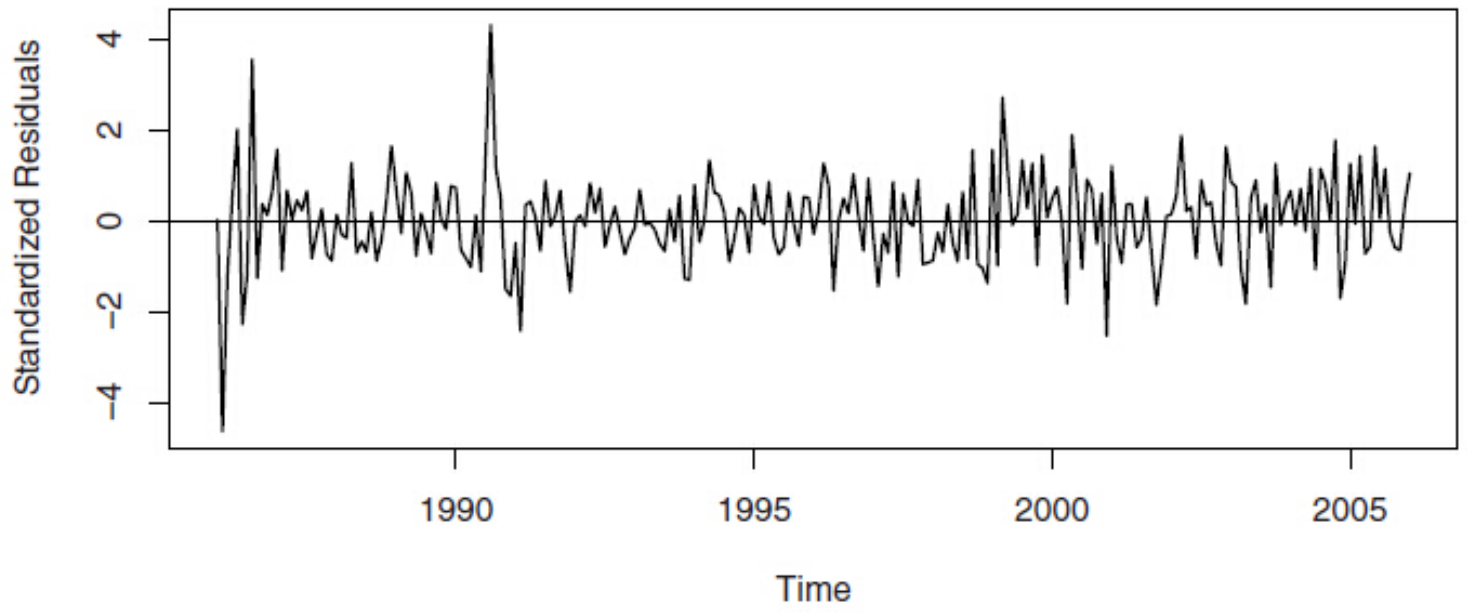


The model of $\texttt{hare}^{0.5}$ built in Exhibit 7.8 is an AR(3) model with ϕ_2 term dropped:

$$\sqrt{Y_t} = 3.483 + 0.919\sqrt{Y_{t-1}} - 0.5313\sqrt{Y_{t-3}} + e_t$$

From the standardized residual plot, the variation looks not quite consistent. The model is not so ideal. (The seemingly large negative standardized residuals are not outliers according to the Bonferroni outlier criterion with critical values ± 3.15.)

Exhibit 8.3 Standardized Residuals from Log Oil Price IMA(1,1) Model



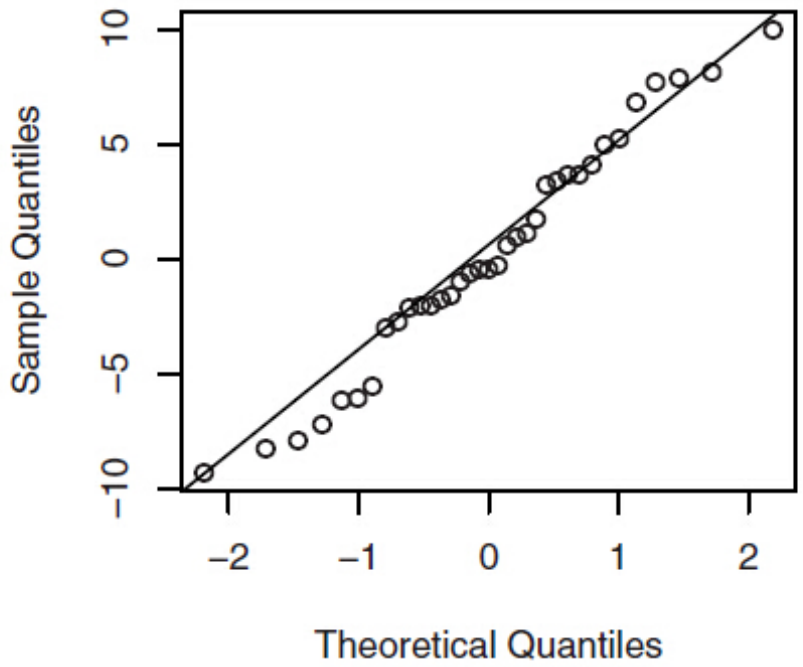

```
> data(oil.price)
> m1.oil=arima(log(oil.price),order=c(0,1,1))
> plot(rstandard(m1.oil),ylab='Standardized residuals',type='l')
> abline(h=0)
```

The standardized residual plot looks fine, except for a few spikes. At least two or three residuals have magnitude greater than 3, which is rare for a standard white noise process. We should investigate the other factors influencing the change of oil prices. (The Bonferroni critical values with $n = 241$ and $\alpha = 0.05$ are ± 3.71, so the outliers do appear to be real.)

8.1.2 Normality of the Residuals

We use quantile-quantile plots and Shapiro-Wilk normality test on standardized residuals to assess normality.

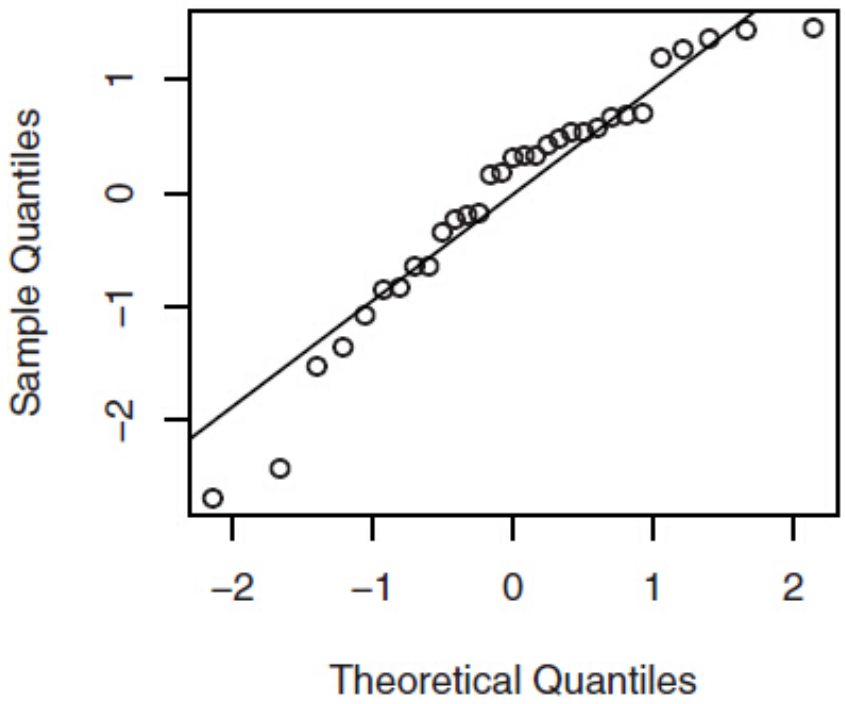
Exhibit 8.4 Quantile-Quantile Plot: Residuals from AR(1) Color Model




```
> win.graph(width=2.5,height=2.5,pointsize=8)
> qqnorm(residuals(m1.color)); qqline(residuals(m1.color))
```

The QQ plot and the Shapiro-Wilk normality test ($W = 0.9754$, p-value$= 0.6057$) would not lead us to reject normality of the error terms in this model.

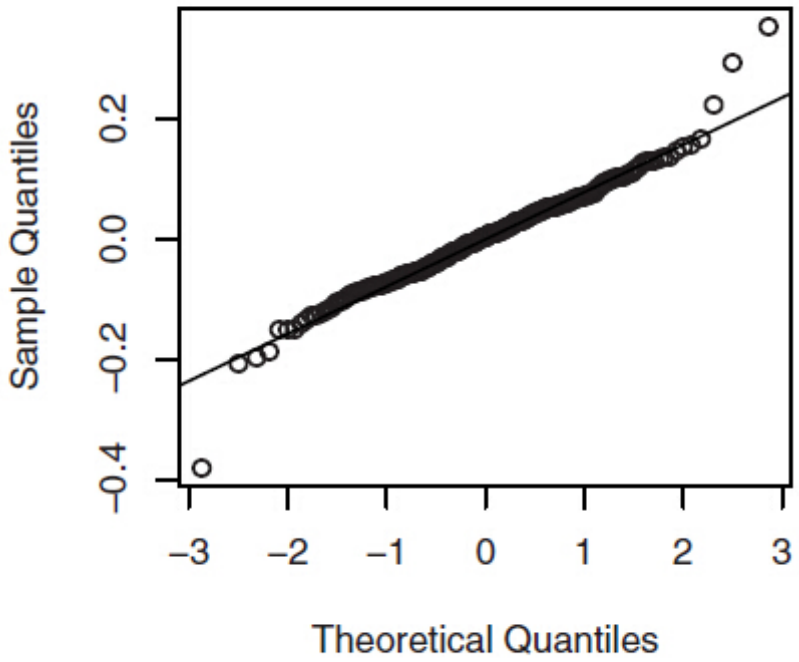
Exhibit 8.5 Quantile-Quantile Plot: Residuals from AR(3) for Hare




```
> qqnorm(residuals(m1.hare)); qqline(residuals(m1.hare))
```

The extreme values look suspect. However, the sample is small ($n = 31$) and the Bonferroni criteria do not indicate outliers. Shapiro-Wilk Test does not reject null hypothesis of normality.

Exhibit 8.6 Quantile-Quantile Plot: Residuals from IMA(1,1) Model for Oil




```
> qqnorm(residuals(m1.oil)); qqline(residuals(m1.oil))
```

The outliers are quite prominent. Shapiro-Wilk Test would reject the null hypothesis.

8.1.3 Autocorrelation of the Residuals

For true white noise with large n, the sample autocorrelations r_k are approximately normally distributed with zero means. We have seen that

$$Var(r_k) \approx \frac{1}{n} \text{ and } Corr(r_k, r_j) \approx 0 \text{ for } k \neq j \qquad (6.1.3)$$

For the sample autocorrelations $\hat{r}_k$ of the residuals in a correctly specified model, however, $\mathrm{Var}(\hat{\mathrm{r}}_\mathrm{k})$ has slightly different behaviors:

- for small lags k and j, $\mathrm{Var}(\hat{\mathrm{r}}_\mathrm{k})$ can be substantially less than $1/n$ and the estimates $\mathrm{Var}(\hat{\mathrm{r}}_\mathrm{k})$ and $\mathrm{Var}(\hat{\mathrm{r}}_\mathrm{j})$ can be highly correlated.
- For larger lags, the approximate variance $1/n$ does apply; furthermore, $\mathrm{Var}(\hat{\mathrm{r}}_\mathrm{k})$ and $\mathrm{Var}(\hat{\mathrm{r}}_\mathrm{j})$ are approximately uncorrelated.

For an AR(1) Model, for large n, $\mathrm{Var}\left(\hat{r}_1\right) \approx \frac{\phi^2}{n}$ and $\mathrm{Var}\left(\hat{r}_k\right) \approx \frac{1}{n}$ for $k > 1$. Exhibit 8.7 shows the approximations for Residual Autocorrelations in AR(1) Models.

Exhibit 8.7 Approximations for Residual Autocorrelations in AR(1) Models

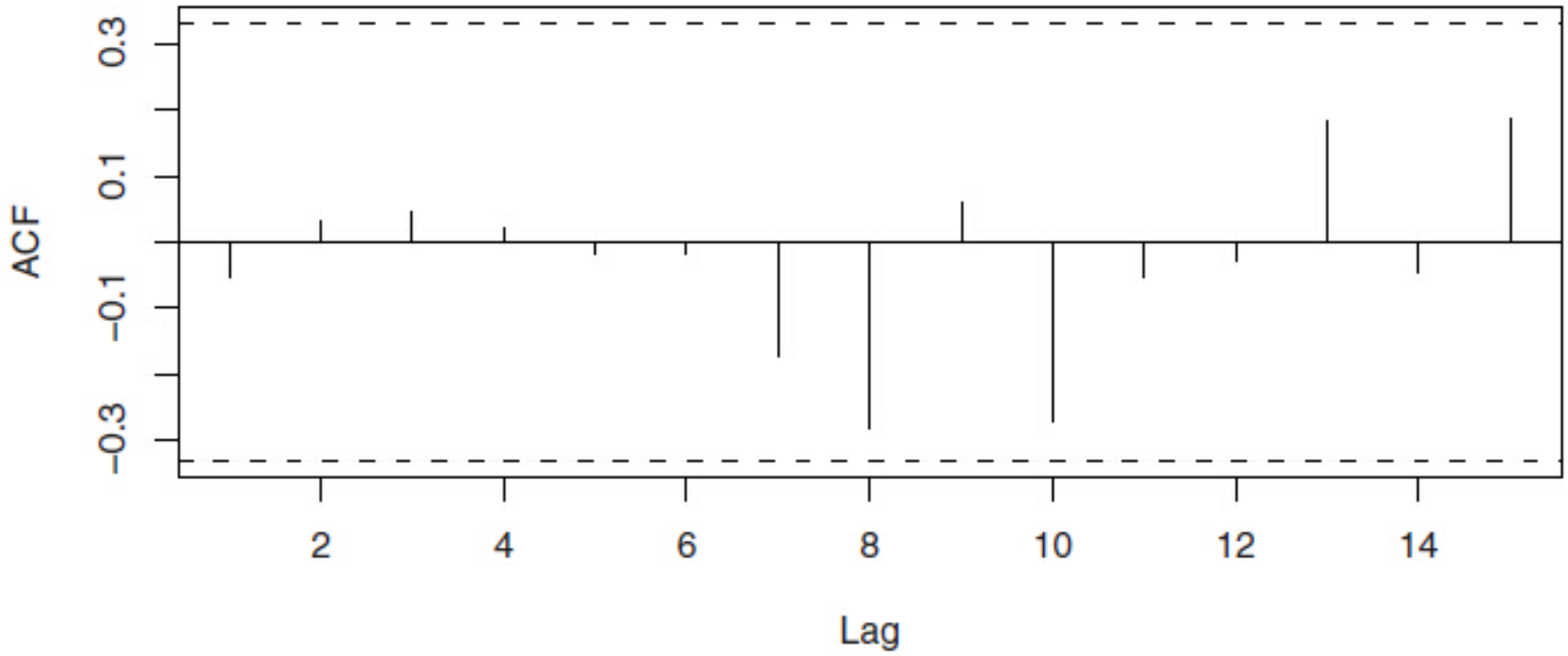
ϕ	0.3	0.5	0.7	0.9	ϕ	0.3	0.5	0.7	0.9
k	Standard deviation of $\hat{r}_k$ times $\sqrt{n}$					Correlation $\hat{r}_1$ with $\hat{r}_k$			
1	0.30	0.50	0.70	0.90		1.00	1.00	1.00	1.00
2	0.96	0.90	0.87	0.92		–0.95	–0.83	–0.59	–0.21
3	1.00	0.98	0.94	0.94		–0.27	–0.38	–0.38	–0.18
4	1.00	0.99	0.97	0.95		–0.08	–0.19	–0.26	–0.16
5	1.00	1.00	0.99	0.96		–0.02	–0.09	–0.18	–0.14
6	1.00	1.00	0.99	0.97		–0.01	–0.05	–0.12	–0.13
7	1.00	1.00	1.00	0.97		–0.00	–0.02	–0.09	–0.12
8	1.00	1.00	1.00	0.98		–0.00	–0.01	–0.06	–0.10
9	1.00	1.00	1.00	0.99		–0.00	–0.00	–0.03	–0.08

If we apply these results to the AR(1) model that was estimated for the industrial color property time series with $\hat{\phi} = 0.57$ and $n = 35$, we obtain the results shown in Exhibit 8.8.

Exhibit 8.8 Approximate Standard Deviations of Residual ACF values

Lag k	1	2	3	4	5	> 5
$\sqrt{Var(\hat{r}_k)}$	0.096	0.149	0.163	0.167	0.168	0.169

Exhibit 8.9 Sample ACF of Residuals from AR(1) Model for Color




```
> win.graph(width=4.875,height=3,pointsize=8)
> acf(residuals(m1.color))
```

There is no evidence of autocorrelation in the residuals of this model.

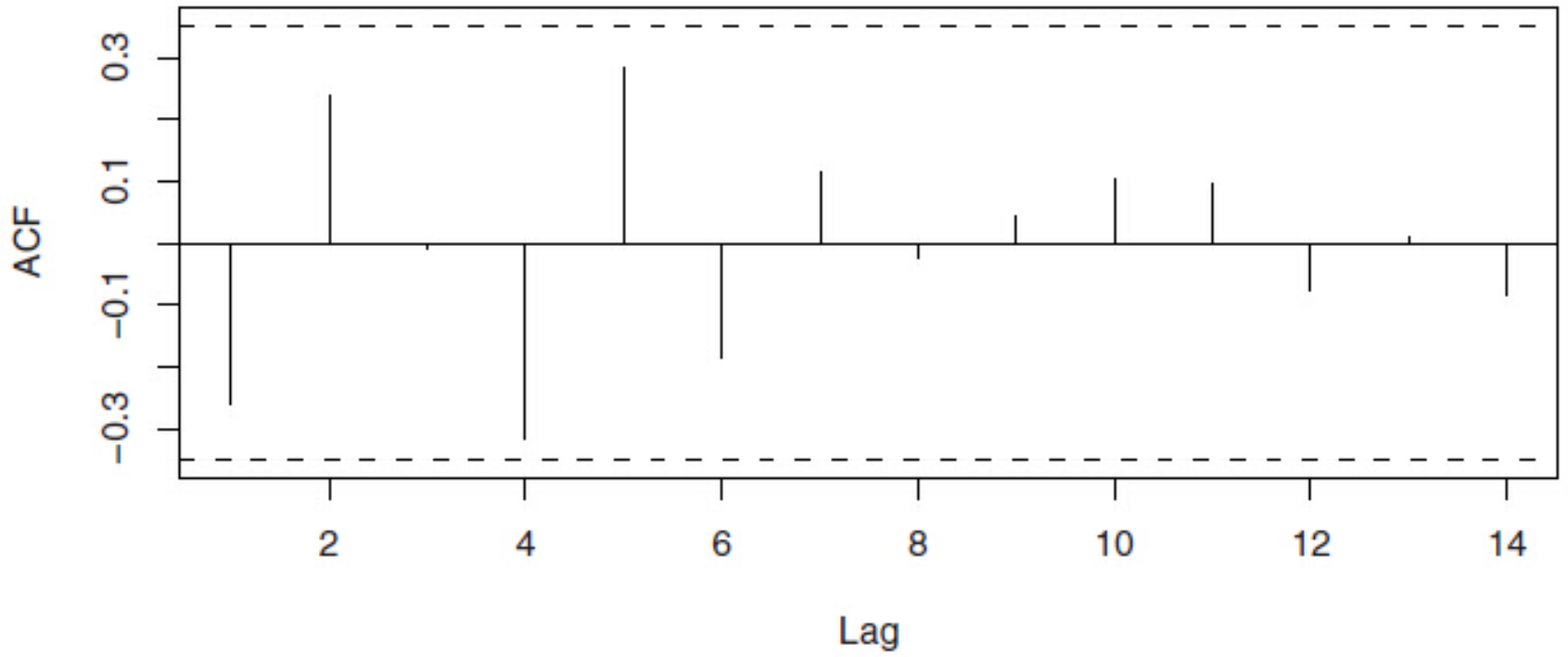
For an AR(2) model, we have

$$\mathrm{Var}\left(\hat{\mathrm{r}}_1\right) \approx \frac{\phi_2^2}{\mathrm{n}}, \qquad \mathrm{Var}\left(\hat{\mathrm{r}}_2\right) \approx \frac{\phi_2^2 + \phi_1^2(1+\phi_2)^2}{\mathrm{n}}$$

If the AR(2) parameters are not too close to the stationarity boundary, then

$$\mathrm{Var}\left(\hat{\mathrm{r}}_\mathrm{k}\right) \approx \frac{1}{\mathrm{n}} \quad \text{for } \mathrm{k} \geq 3.$$

Exhibit 8.10 Sample ACF of Residuals from AR(2) Model for Hare




```
> acf(residuals(arima(sqrt(hare),order=c(2,0,0))))
```

The graph does not show statistically significant evidence of nonzero autocorrelation in the residuals.

With monthly data, we may pay special attention to the residuals at lags 12, 24, and so forth. With quarterly series, lags 4, 8, and so forth would merit special attention. Chapter 10 contains examples of these ideas.

The residuals for MA models have similar results.

8.1.4 The Ljung-Box Test

To study the independence of the residuals, we may takes into account the magnitudes of all residual correlations $\{\hat{r}_k\}$ as a group. The **Ljung-Box** test statistic is

$$Q_* = n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \cdots + \frac{\hat{r}_K^2}{n-K}\right)$$

where K is selected large enough such that the Ψ-weights are negligible for $j > K$. We would reject the ARMA(p,q) model if the observed value of Q_* exceeded an appropriate critical value in a chi-square distribution with $K - p - q$ degrees of freedom.

Exhibit 8.11 Residual Autocorrelation Values from AR(1) Model for Color

Lag k	1	2	3	4	5	6
Residual ACF	–0.051	0.032	0.047	0.021	–0.017	–0.019

```
> acf(residuals(m1.color),plot=F)$acf
> signif(acf(residuals(m1.color),plot=F)$acf[1:6],2)
> # display the first 6 acf values to 2 significant digits
```

The Ljung-Box test statistic with $K = 6$ is equal to

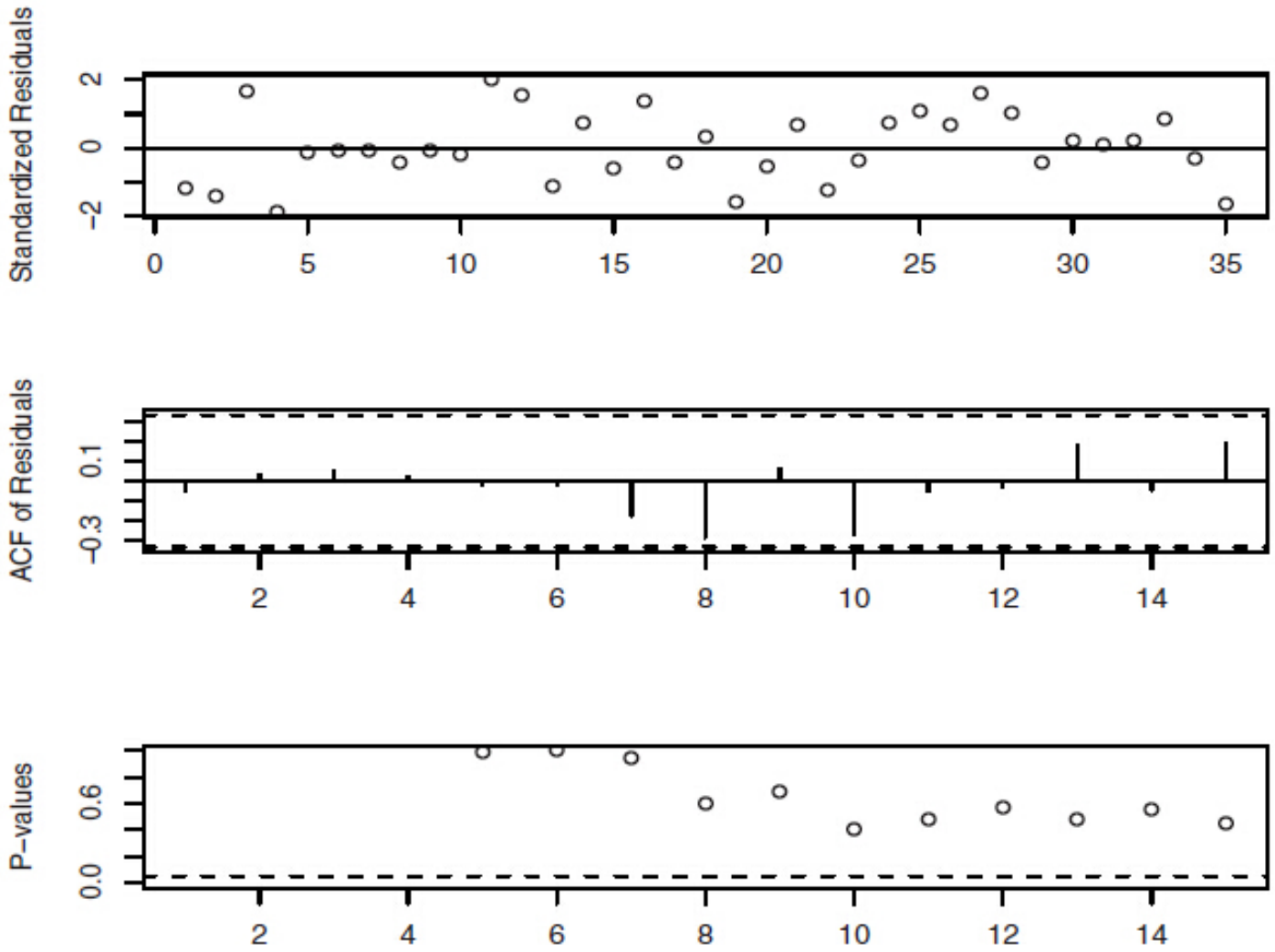
$$Q_* = 35(35+2)\left(\frac{(-0.051)^2}{35-1} + \frac{(0.032)^2}{35-2} + \frac{(0.047)^2}{35-3} + \frac{(0.021)^2}{35-4} + \frac{(-0.017)^2}{35-5} + \frac{(-0.019)^2}{35-6}\right) \approx 0.28$$

This is referred to a chi-square distribution with $6 - 1 = 5$ degrees of freedom. This leads to a p-value of 0.998, so we have no evidence to reject the null hypothesis that the error terms are uncorrelated.

The Ljung-Box test statistic can be calculated by the command:

```
LB.test(m1.color, lag=6)
```

Exhibit 8.12 Diagnostic Display for the AR(1) Model of Color Property




```
> win.graph(width=4.875,height=4.5)
> tsdiag(m1.color,gof=15,omit.initial=F)
```

Exhibit 8.12 (cont) The command `tsdiag` displays three of our diagnostic tools in one display—a sequence plot of the standardized residuals, the sample ACF of the residuals, and p-values for the Ljung-Box test statistic for a whole range of values of K from 5 to 15. Everything looks very good. The estimated AR(1) model seems to be capturing the dependence structure of the color property time series quite well. The `runs` test also do not reject the independence of the residuals.

```
tsdiag(m1.color,gof=15,omit.initial=F)
```

The `tsdiag` function in the `TSA` package has been modified from that in the `stats` package of R. It performs model diagnostics on a fitted model.

1. The argument `gof` specifies the maximum number of lags in the `acf` function used in the model diagnostics. The maximum lag should be set sufficiently large, so that the original time series has negligible autocorrelation, and thus the Ljung-Box (portmanteau) test statistic can be performed.
2. Setting the argument `omit.initial=T` omits the few initial residuals from the analysis. This option is especially useful for checking seasonal models where the initial residuals are close to zero by construction and including them may skew the model diagnostics. In the example, the `omit.initial` argument is set to be `F` so that the diagnostics are done with all residuals.

Ex. [Exercise 8.3, TS-ch8.R] *Based on a series of length* $n = 200$, *we fit an AR(2) model and obtain residual autocorrelations of*

$$\hat{r}_1 = 0.13, \quad \hat{r}_2 = 0.13, \quad \text{and } \hat{r}_3 = 0.12.$$

If $\hat{\phi}_1 = 1.1$ *and* $\hat{\phi}_2 = -0.8$, *do these residual autocorrelations support the AR(2) specification? Individually? Jointly?*

Exercise 8.3 Solution: For an AR(2) model, we first use (8.1.8), (8.1.9), (8.1.10):

$$\sqrt{\mathrm{Var}\left(\hat{\mathrm{r}}_1\right)} \approx \sqrt{\frac{\phi_2^2}{n}} = \sqrt{\frac{(-0.8)^2}{200}} = 0.057, \quad \text{so } \hat{r}_1 = 0.13 \text{ is too large;}$$

$$\sqrt{\mathrm{Var}\left(\hat{\mathrm{r}}_2\right)} \approx \sqrt{\frac{\phi_2^2 + \phi_1^2(1+\phi_2)^2}{n}} = 0.059, \quad \text{so } \hat{r}_2 = 0.13 \text{ is too large;}$$

$$\sqrt{\mathrm{Var}\left(\hat{\mathrm{r}}_k\right)} \approx \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{200}} = 0.071 \quad \text{for } k \geq 3, \quad \text{so } \hat{r}_3 = 0.12 \text{ is fine;}$$

The Ljung-Box statistic:

$$Q_* = n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \frac{\hat{r}_3^2}{n-3}\right) = 9.83,$$

if the AR(2) model is correct, then Q_* is approximately a chi-square distribution with $3-2-0=1$ degree of freedom. Check the commands of chi-square distribution by "`?Chisquare`". Here we use "`1-pchisq(9.83, df=1)`" to see that $Prob[Q_* > 9.83] = 0.0017$. Hence the residual autocorrelations are (jointly) too large to support the AR(2) model.

Ex. *(Exercise 8.7, TS-ch8.R) Fit an AR(3) model by maximum likelihood to the square root of the hare abundance series (filename* `hare`*).*

1. *Plot the sample ACF of the residuals. Comment on the size of the correlations.*
2. *Calculate the Ljung-Box statistic summing to $K = 9$. Does this statistic support the AR(3) specification?*
3. *Perform a runs test on the residuals and comment on the results.*
4. *Display the quantile-quantile normal plot of the residuals. Comment on the plot.*
5. *Perform the Shapiro-Wilk test of normality on the residuals.*

Overfitting and Parameter Redundancy

Another basic diagnostic tool is that of **overfitting**. After fitting a likely adequate model, we fit a slightly more general model and compare it with the original model. The original model would be confirmed if:

1. the estimate of the additional parameter is not significantly different from zero, and
2. the estimates for the parameters in common do not change significantly from their original estimates.

We may compare the other statistics, such as log-likelihood and AIC, as well.

Ex. *We have specified, fitted, and examined the residuals of an AR(1) model for the industrial color property time series. Let us compare it with slightly more general models: AR(2) and ARMA(1,1) models.*

Exhibit 8.13 AR(1) Model Results for the Color Property Series

Coefficients:†	ar1	Intercept‡
	0.5705	74.3293
s.e.	0.1435	1.9151

sigma^2 estimated as 24.83: log-likelihood = -106.07, AIC = 216.15

† `m1.color` # R code to obtain table

‡ Recall that the intercept here is the estimate of the process mean μ—not θ_0.

Exhibit 8.14 AR(2) Model Results for the Color Property Series

Coefficients:	ar1	ar2	Intercept
	0.5173	0.1005	74.1551
s.e.	0.1717	0.1815	2.1463

sigma^2 estimated as 24.6: log-likelihood = -105.92, AIC = 217.84

```
> arima(color,order=c(2,0,0))
```

(cont.) Here are some observations:

1. In Exhibit 8.14, the estimate of ϕ_2 is not statistically different from zero;
2. two estimates of ϕ_1 and μ are quite close;
3. while the AR(2) model has a slightly larger log-likelihood value, the AR(1) fit has a smaller AIC value.

Therefore, the simpler AR(1) model is better.

Next we compare the AR(1) model with the ARMA(1,1) model.

Ex.

Exhibit 8.15 Overfit of an ARMA(1,1) Model for the Color Series

Coefficients:	ar1	ma1	Intercept
	0.6721	–0.1467	74.1730
s.e.	0.2147	0.2742	2.1357

sigma^2 estimated as 24.63: log-likelihood = -105.94, AIC = 219.88

```
> arima(color,order=c(1,0,1))
```

Observations:

1. *the new parameter θ is not significantly different from zero;*
2. *the estimate of ϕ_1 and μ are not significantly different from the estimate in Exhibit 8.13;*
3. *the AIC is greater than that of the AR(1) fit.*

Again the AR(1) model looks better.

When generalizing ARMA models, we must be aware of the problem of **parameter redundancy**.

Define the **backshift operation** B on any process $\{Y_t\}$ by $B(Y_t) = Y_{t-1}$. For example, an ARMA(1,1) model: $Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$, may be expressed as $Y_t - \phi Y_{t-1} = e_t - \theta e_{t-1}$, which is $(1 - \phi B)Y_t = (1 - \theta B)e_t$.

If we have a correct ARMA model: $\phi(B)Y_t = \theta(B)e_t$, then $(1 - cB)\phi(B)Y_t = (1 - cB)\theta(B)e_t$ is also a correct model. This creates parameter redundancy. To get unique ARMA model, we must cancel any common factors in the AR and MA characteristic polynomials.

To avoid parameter redundancy:

1. Try simple models first.
2. When overfitting, do not increase the the AR and MA orders simultaneously.
3. Extend the model in directions suggested by the analysis (e.g. ACF, PACF) of the residuals.

Ex.

Exhibit 8.16 Overfitted ARMA(2,1) Model for the Color Property Series

Coefficients:	ar1	ar2	ma1	Intercept
	0.2189	0.2735	0.3036	74.1653
s.e.	2.0056	1.1376	2.0650	2.1121

sigma^2 estimated as 24.58: log-likelihood = −105.91, AIC = 219.82

```
> arima(color,order=c(2,0,1))
```

The estimates of ϕ_1, ϕ_2, and θ in this ARMA(2,1) fit are very different from those in AR(1) fit, and they are not significantly different from zero.

Ex. *(TS-ch8.R) We simulate an ARMA(2,1) series with* $\phi_1 = -0.3, \phi_2 = 0.4, \theta_1 = -0.5$, *and* $n = 1000$. *Then we specify the model, estimate the parameters, and analyze the residuals.*