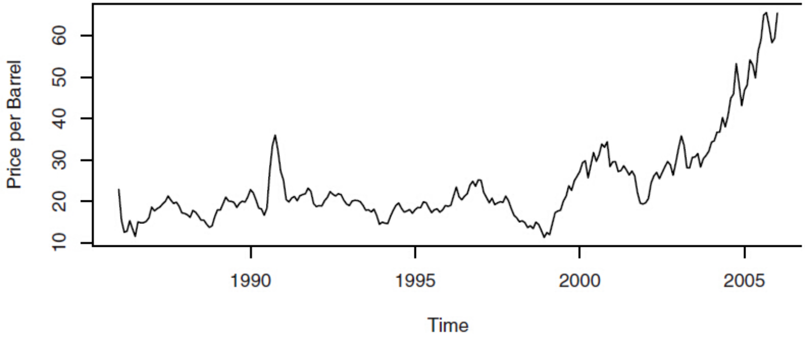
Ch 5. Models for Nonstationary Time Series

Time Series Analysis

We have studied some deterministic and some stationary trend models. However, many time series data cannot be modeled in either way.

Exhibit 5.1 Monthly Price of Oil: January 1986–January 2006



Ex. The data set `oil.price` displays an increasing variation from the plot. No stationary model fits the data (neither does a deterministic trend model.)

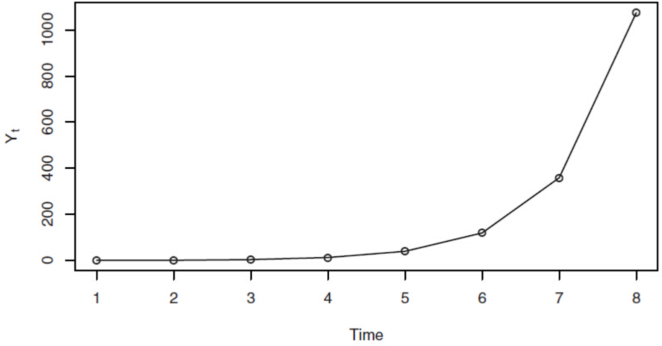
5.1 Stationarity Through Differencing

The stationarity condition of an AR(1) model: $Y_t = \phi Y_{t-1} + e_t$ is $|\phi| < 1$. If $|\phi| \geq 1$, we will get nonstationary models.

Exhibit 5.2 **Simulation of the Explosive "AR(1) Model"** $Y_t = 3Y_{t-1} + e_t$

t	1	2	3	4	5	6	7	8
e_t	0.63	–1.25	1.80	1.51	1.56	0.62	0.64	–0.98
Y_t	0.63	0.64	3.72	12.67	39.57	119.33	358.63	1074.91

Exhibit 5.3 **An Explosive "AR(1)" Series**



Ex. We could simulate another Explosive "AR(1) Model" $Y_t = 3Y_{t-1} + e_t$. See `TS-ch5.R`.

Taking difference is one possible way to obtain stationary model.

Def. For a process $\{Y_t\}$, we define **the first difference of** Y_t as

$$\nabla Y_t = Y_t - Y_{t-1}.$$

Ex. For the random walk model: $Y_t = Y_{t-1} + e_t$, $\nabla Y_t = e_t$ is a stationary process.

In may cases, time series can be thought of being composed of a nonstationary trend component and a zero-mean stationary component.

Ex. Some assumptions may lead to stationary second-difference models.

1. If we assume that $Y_t = M_t + X_t$, where M_t is linear in time over three consecutive time points, we may predict M_t at middle time point t by choosing $\beta_{0,t}$ and $\beta_{1,t}$ to minimize

$$\sum_{j=-1}^{1} (Y_{t-j} - (\beta_{0,t} + j\beta_{1,t}))^2.$$

The solution is $\hat{M}_t = \frac{Y_{t-1}+Y_t+Y_{t+1}}{3}$. So the detrended series is

$$\hat{X}_t = Y_t - \hat{M}_t = Y_t - \frac{Y_{t-1} + Y_t + Y_{t+1}}{3}$$
$$= -\frac{1}{3}\left[(Y_{t+1} - Y_t) - (Y_t - Y_{t-1})\right] = -\frac{1}{3}\nabla^2 Y_t$$

2. In the following model,

$$Y_t = M_t + e_t, \qquad M_t = M_{t-1} + W_t, \qquad W_t = W_{t-1} + \epsilon_t,$$

we see that $\nabla^2 Y_t$ is stationary:

$$\begin{aligned}\nabla Y_t &= \nabla M_t + \nabla e_t = W_t + \nabla e_t,\\ \nabla^2 Y_t &= \nabla W_t + \nabla^2 e_t = \epsilon_t + e_t - 2e_{t-1} + e_{t-2}\end{aligned}$$

5.2 ARIMA Models

Def. [Integrated autoregressive moving average model] A process $\{Y_t\}$ is said to be ARIMA(p,d,q) if the dth difference $W_t = \nabla^d Y_t$ is a stationary ARMA(p,q) process.

Ex. Suppose $\{Y_t\} \sim ARIMA(p, 1, q)$. Let $W_t = Y_t - Y_{t-1}$. Then $\{W_t\} \sim ARMA(p, q)$. Suppose

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Then $\{Y_t\}$ satisfies that

$$\begin{aligned} Y_t - Y_{t-1} \quad &= \quad \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \cdots + \phi_p(Y_{t-p} - Y_{t-p-1}) \\ &\qquad + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \end{aligned}$$

Combining the terms, we get the **difference equation form** of Y_t. It looks like a ARMA(p+1,q) process. However, the AR characteristic polynomial of Y_t would be $(1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p)(1 - x)$, which has a root 1. So the "ARMA(p+1,q) process" is not stationary!

In general, if $\{Y_t\} \sim ARIMA(p, d, q)$ and $W_t = \nabla^d Y$ has the AR characteristic polynomial $\phi(x)$. Then Y has a AR characteristic polynomial $\phi(x)(1-x)^d$. We can use this to determine the types of ARIMA models.

The nonstationary Y_t may be represented as a sum of stationary W_t starts at certain $t = -m < 1$, and assuming $Y_t = 0$ for $t < -m$. This expression is good to find covariance properties.

Ex. For a ARIRM(p,1,q) model, $W_t = Y_t - Y_{t-1}$ is stationary, and the nonstationary Y_t can be expressed as

$$Y_t = \sum_{j=-m}^{t} W_j.$$

Similarly for the general $ARIMA(p, d, q)$ models.

Def. Two special families of nonstationary processes:

- IMA(d,q)=ARIMA(0,d,q): the process have no autoregressive terms;
- ARI(p,d)=ARIMA(p,d,0): the process has no moving average terms.

Ex. [HW 5.10] Nonstationary ARIMA series can be simulated by first simulating the corresponding stationary ARMA series and then "integrating" it (really partially summing it). Use statistical software to simulate a variety of IMA(1,1) and IMA(2,2) series with a variety of parameter values. Note any stochastic "trends" in the simulated series.

```
# Here is an example of IMA(1,1) model
S1=arima.sim(model=list(order=c(0,1,1),
ma=-0.7),n=30)
```

5.2.1 The IMA(1,1) Model: Many economic time series can be modeled as IMA(1,1). Difference equation form:

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1}.$$

It can be represented as:

$$\begin{aligned} Y_t &= \sum_{j=-m}^{t} (Y_j - Y_{j-1}) = \sum_{j=-m}^{t} (e_j - \theta e_{j-1}) \\ &= e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1}. \end{aligned}$$

The coefficients of the past white noise do not die out. Explicit computation shows that

$$\begin{aligned} \mathrm{Var}(Y_t) &= [1 + \theta^2 + (1-\theta)^2(t+m)]\sigma_e^2, \\ \mathrm{Corr}(Y_t, Y_{t-k}) &\approx \sqrt{1 - \frac{k}{m+t}}. \end{aligned}$$

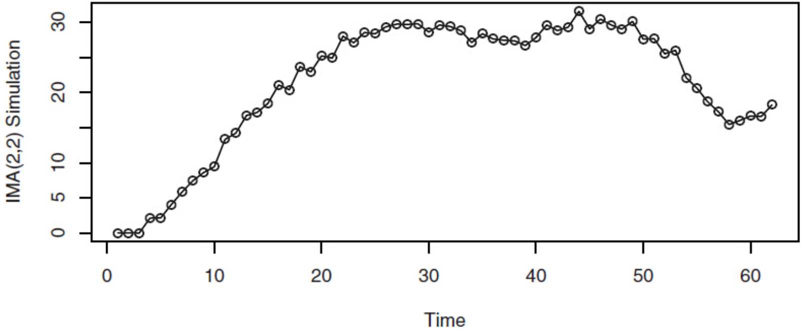
Like the random walk process, $\mathrm{Corr}(Y_t, Y_{t-k}) \approx 1$ for large $m + t$ and moderate k.

5.2.2 The IMA(2,2) Model: Difference equation form:

$$\nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}.$$

We may express Y_t as a sum of white noise terms:
$Y_t = \sum_{j=-m-2}^{t} \Psi_j e_j$, where Ψ_j is a linear function of j. We will see that $\mathrm{Cov}\,(\mathrm{Y_t})$ increase rapidly with t, and $\mathrm{Corr}\,(\mathrm{Y_t}, \mathrm{Y_{t-k}}) \approx 1$ for large $t + m$ and moderate k.

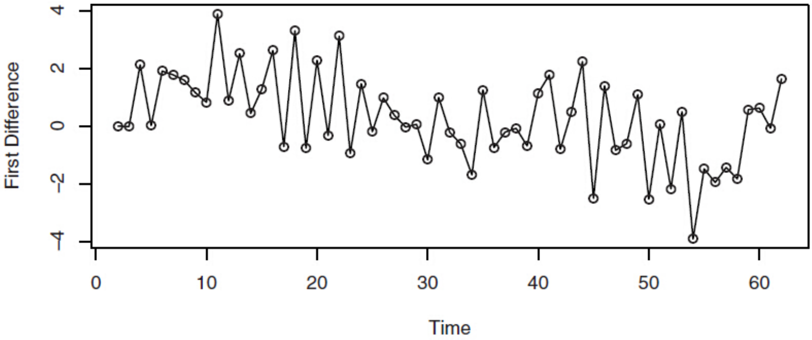
Exhibit 5.5 **Simulation of an IMA(2,2) Series with $\theta_1 = 1$ and $\theta_2 = -0.6$**




```
> data(ima22.s)
> plot(ima22.s,ylab='IMA(2,2) Simulation',type='o')
```

The data set `ima22.s` is a simulation of an IMA(2,2) Series with $\theta_1 = 1$ and $\theta_2 = -0.6$. We can see the increasing variance and the strong positive neighboring correlations.

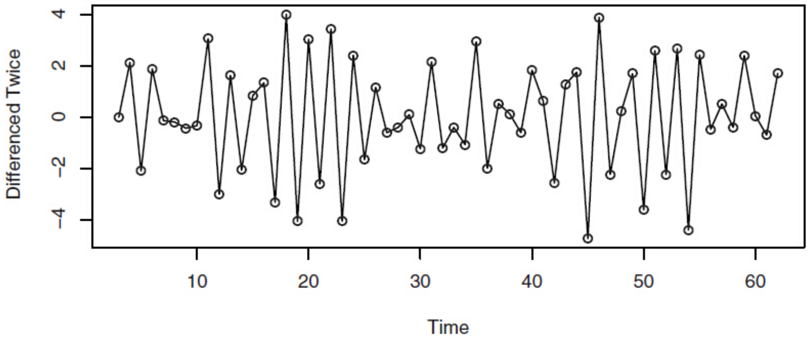
Exhibit 5.6 First Difference of the Simulated IMA(2,2) Series




```
> plot(diff(ima22.s),ylab='First Difference',type='o')
```

The first difference of the simulated IMA(2,2) series `ima22.s`.
$\nabla Y_t \sim IMA(1,2)$ seems nonstationary.

Exhibit 5.7 Second Difference of the Simulated IMA(2,2) Series




```
> plot(diff(ima22.s,difference=2),ylab='Differenced
   Twice',type='o')
```

The second difference of the simulated IMA(2,2) series `ima22.s`. $\nabla^2 Y_t \sim MA(2)$ is stationary. The plot seems consistent with the theoretical ACF $\rho_1 = -0.678$ and $\rho_2 = 0.254$.

5.2.3 The ARI(1,1) Model: The difference equation:
$Y_t - Y_{t-1} = \phi(Y_{t-1} - Y_{t-2}) + e_t$, or

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t, \qquad |\phi| < 1.$$

Theorem 1

In general, for an ARIMA(p,d,q) model $\{Y_t\}$, the Ψ-weights can be calculated by equating the following identity:

$$\begin{aligned} & (1-\phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p)(1-x)^d(1+\Psi_1 x + \Psi_2 x^2 + \cdots) \\ = \; & 1 - \theta_1 x - \theta_2 x^2 - \cdots - \theta_q x^q. \end{aligned}$$

For the ARI(1,1) model, we have

$$\begin{aligned} \Psi_1 &= 1+\phi \\ \Psi_2 &= (1+\phi)\Psi_1 - \phi \\ \Psi_k &= (1+\phi)\Psi_{k-1} - \phi\Psi_{k-2}, \quad k \geq 2. \end{aligned}$$

We can solve that $\Psi_k = \frac{1-\phi^{k+1}}{1-\phi}$ for $k \geq 1$.

5.3 Constant Terms in ARIMA Models

If $\{Y_t\} \sim ARIMA(p, d, q)$, then $\{W_t = \nabla^d Y_t\} \sim ARMA(p, q)$. When W_t has constant mean $\mu \neq 0$, we can either model it as:

$$W_t - \mu = \sum_{i=1}^{p} \phi_i (W_{t-i} - \mu) + e_t - \sum_{j=1}^{q} \theta_j e_{t-j}$$

or introduce a constant θ_0 into the model.

$$W_t = \theta_0 + \sum_{i=1}^{p} \phi_i W_{t-i} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j}.$$

Taking expected value, we get $\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$.

In general, when $\{Y_t\} \sim ARIMA(p, d, q)$ with $E(\nabla^d Y_t) \neq 0$, we have $Y_t = Y_t' + \mu_t$ where μ_t is a deterministic polynomial of degree d, and $Y_t' \sim ARIMA(p, d, q)$ has zero mean.

5.5 Other Transformations

There are some other common transformations to achieve stationarity: logarithm, percentage change, and power transformations.

- (logarithm transformation) Suppose $Y_t > 0$ for all t, and

$$E(Y_t) = \mu_t, \qquad \sqrt{\mathrm{Var}\,(\mathrm{Y_t})} = \mu_t\sigma.$$

By Taylor expansion,

$$\log Y_t = \log \mu_t + \log\left(1 + \frac{Y_t - \mu_t}{\mu_t}\right) \approx \log \mu_t + \frac{Y_t - \mu_t}{\mu_t}.$$

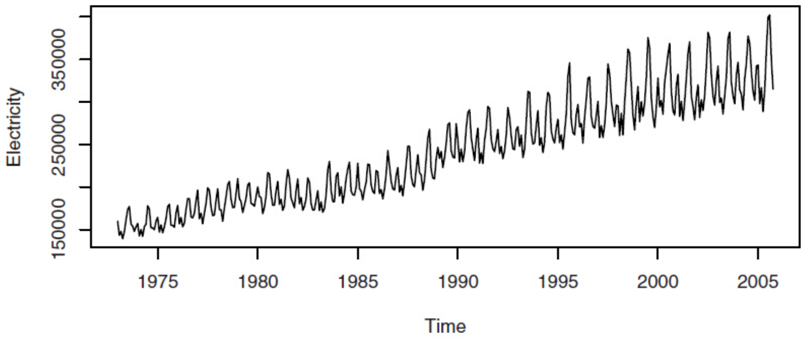
So $E(\log Y_t) \approx \log \mu_t$. Similarly, we get $\mathrm{Var}\,(\log \mathrm{Y_t}) \approx \sigma^2$.

- Assume that $Y_t = (1 + X_t)Y_{t-1}$. Then $100X_t$ is the percentage change. Suppose $\{Y_t\}$ tends to have relative stable percentage change. When $|X_t| < 0.2$, we may use logarithm transformation for $\{Y_t\}$ as well:

$$\nabla[\log(Y_t)] = \log(Y_t/Y_{t-1}) = \log(1 + X_t) \approx X_t,$$

which may be well-modeled by a stationary series.

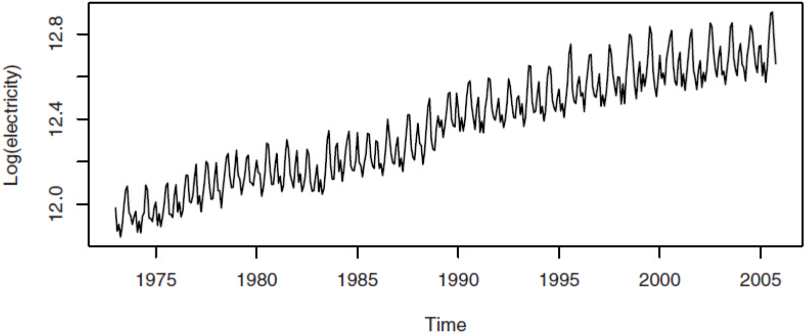
Exhibit 5.8 U.S. Electricity Generated by Month




```
> data(electricity); plot(electricity)
```

The data set `electricity` shows U.S. Electricity Generated by Month. The higher value shows more variation. However, the percentage change might be related stable.

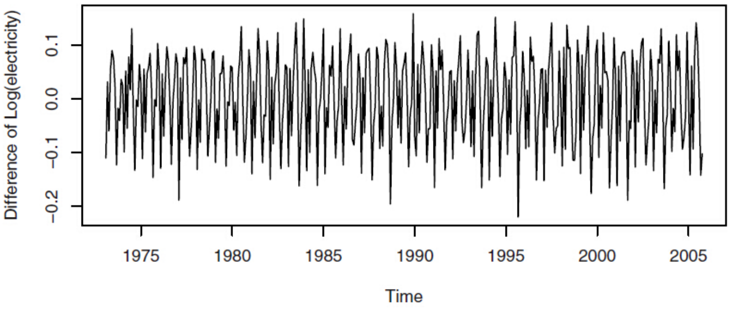
Exhibit 5.9 Time Series Plot of Logarithms of Electricity Values




```
> plot(log(electricity),ylab='Log(electricity)')
```

Above is the time series plot of the logarithms of the electricity values. The variation looks more uniform.

Exhibit 5.10 Difference of Logarithms for Electricity Time Series




```
> plot(diff(log(electricity)),
   ylab='Difference of Log(electricity)')
```

he differences of the logarithms of the electricity values looks like a stationary series. We may compare this plot with the plot of fractional relative change

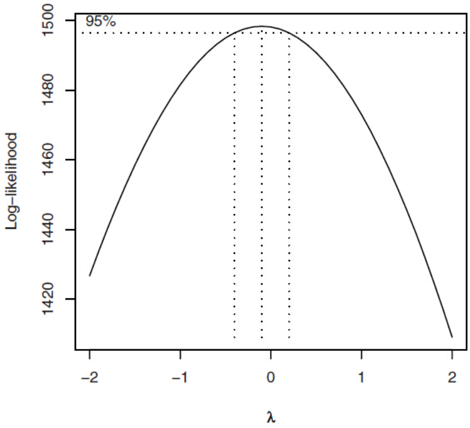
```
elec.frac=na.omit((electricity-zlag(electricity))/zlag(elect
plot(elec.frac,type='o')
plot(y=diff(log(electricity)),x=elec.frac,type='o')
```

- Given a parameter λ, the **power transform** (introduced by Box and Cox) is defined by

$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0, \\ \log x & \text{for } \lambda = 0. \end{cases}$$

  The power transformation applies only to positive data values. If there are negative data, a positive constant may be added to all of the values to make them all positive.
  Appropriate λ may be chosen to transform the data to stabilize the variance and achieve stationarity.

Exhibit 5.11 Log-likelihood versus Lambda




```
> BoxCox.ar(electricity)
```

The R code `BoxCox.ar(electricity)` plots the log-likelihood value for λ. The 95% confidence interval for λ contains the value of $\lambda = 0$ quite near its center and strongly suggests a logarithmic transformation ($\lambda = 0$) for these data. (See discussion of `boxcox` and `BoxCox.ar` commands on Pages 440-441.)