2022 Fall

IE 313 Time Series Analysis

5. Models for Nonstationary Time Series

Yongjae Lee

Department of Industrial Engineering

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Chapter 5. Models for Nonstationary Time Series

Nonstationary Time Series

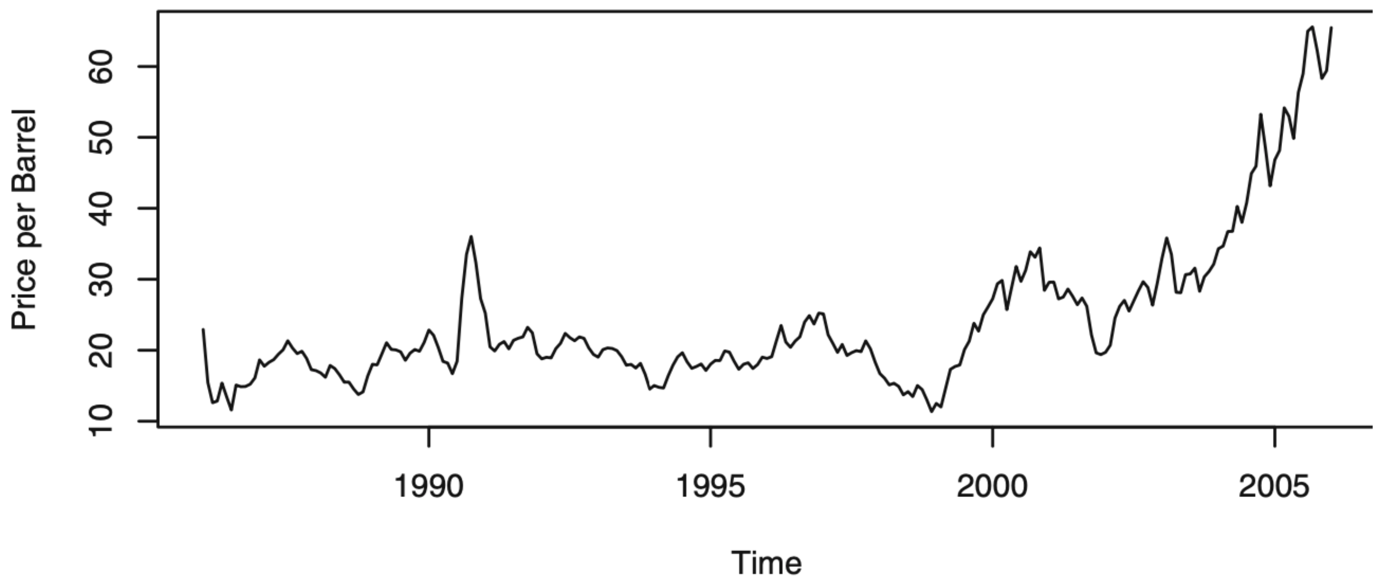
- Any time series without a constant mean over time is nonstationary

$$Y_t = \mu_t + X_t$$

  - Deterministic trend models of the above form were considered in Chapter 3
    - μ_t is a nonconstant mean function
    - X_t is a zero-mean, stationary series
  - But, as we have seen in Chapter 3, finding a deterministic trend is hard and should be careful about it
    - Our observations are often not long enough, while deterministic trends should persist forever

Nonstationary Time Series

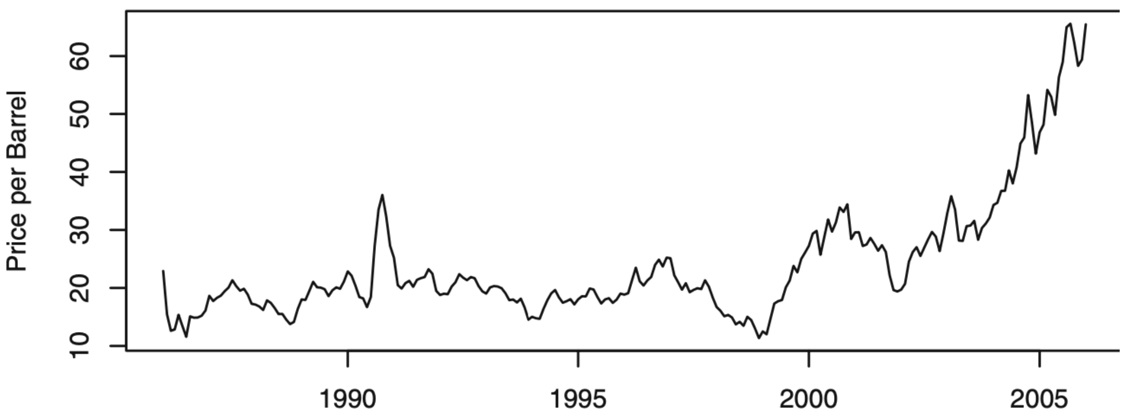
Exhibit 5.1 Monthly Price of Oil: January 1986–January 2006

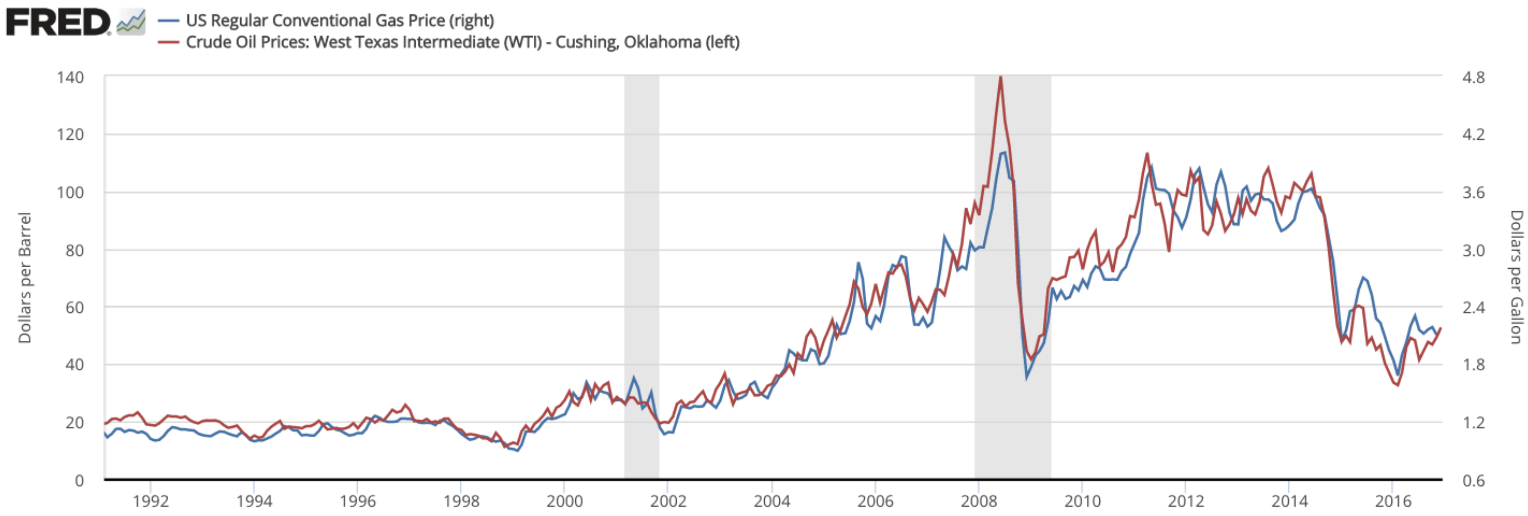


- Can you suggest a good deterministic trend for this data?
  - Actually, no deterministic trend model work for this data
  - Only the nonstationary model with a **stochastic trend** seems reasonable

Nonstationary Time Series

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





Chapter 5.1

Stationarity Through Differencing

Explosive AR

- Consider again the AR(1) model

$$Y_t = \phi Y_{t-1} + e_t$$

  – We have seen that we must have $|\phi| < 1$ for the above model to be stationary

  – What can we say about AR(1) models with $|\phi| > 1$?

Explosive AR

- Let's consider the following model in particular

$$Y_t = 3Y_{t-1} + e_t$$

  – Iterating into the past as we have done before yields

$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + \cdots + 3^{t-1} e_1 + 3^t Y_0$$

  – We see that the influence of distant past values of Y_t and e_t does not die out
    - Rather, the weights applied to Y_0 and e_1 grow exponentially large

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

Explosive AR

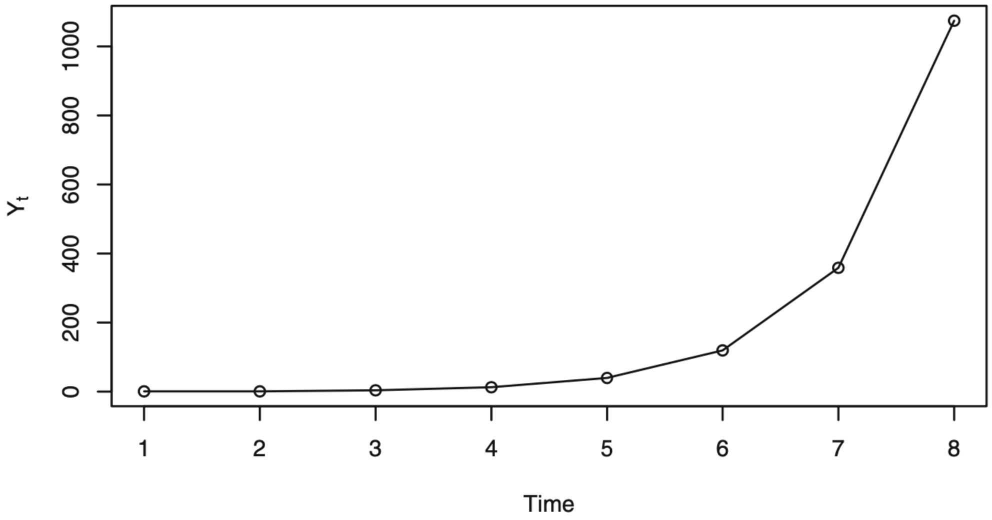
- Let's consider the following model in particular

$$Y_t = 3Y_{t-1} + e_t$$

  – Iterating into the past as we have done before yields

$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + \cdots + 3^{t-1} e_1 + 3^t Y_0$$

Exhibit 5.3 An Explosive "AR(1)" Series



Explosive AR

- Let's consider the following model in particular

$$Y_t = 3Y_{t-1} + e_t$$

  – Iterating into the past as we have done before yields

$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + \cdots + 3^{t-1} e_1 + 3^t Y_0$$

  – The explosive behavior is also reflected in the model's variance and covariance functions

    - $Var(Y_t) = \frac{1}{8}(9^t - 1)\sigma_e^2$
    - $Cov(Y_t, Y_{t-k}) = \frac{3^k}{8}(9^{t-k} - 1)\sigma_e^2$

First difference

- The same general exponential growth or explosive behavior will occur for any $|\phi| > 1$

- A more reasonable type of nonstationarity obtains when $\phi = 1$

$$Y_t = Y_{t-1} + e_t$$

  – We know that this is a random walk process. We can rewrite it as

$$\nabla Y_t = e_t$$

    • Where $\nabla Y_t = Y_t - Y_{t-1}$ is the **first difference** of Y_t

  – We may extend the random walk to be a more general model whose first difference is some stationary process (not just white noise)

First difference

- Some other assumptions can lead to models whose first difference is a stationary process

$$Y_t = M_t + X_t$$

  – Suppose that M_t is a series that is changing only slowly over time (Here, M_t could be either deterministic or stochastic) and X_t is a zero-mean stationary process

  – If we assume that $\boldsymbol{M_t}$ **is approximately constant over every two consecurity time points**, we might estimate (predict) M_t at t by choosing β_0 that achieves

$$\text{minimize} \sum_{j=0}^{1} \left(Y_{t-j} - \beta_{0,t}\right)^2$$

First difference

- It leads to

$$\widehat{M}_t = \frac{1}{2}(Y_t + Y_{t-1})$$

  – And the "detrended" series at time t is then

$$Y_t - \widehat{M}_t = Y_t - \frac{1}{2}(Y_t + Y_{t-1}) = \frac{1}{2}(Y_t - Y_{t-1}) = \frac{1}{2}\nabla Y_t$$

- Hence, assuming that the 'trend (M_t)' is approximately constant over every two consecuritve time points makes **the first difference of Y_t a stationary process**

$Y_t = M_t + X_t$
M_t: approximately constant over every two consecutive time points
X_t: zero-mean stationary process

Estimate M_t by choosing β_0 which minimize $\sum_{j=0}^{1}\left(Y_{t-j} - \beta_{0,t}\right)^2$

First difference

- Another set of assumptions might be that M_t is stocahstic and changes slowly over time governed by a random walk model

$$Y_t = M_t + e_t \qquad \text{with} \quad M_t = M_{t-1} + \varepsilon_t$$

  - Where $\{e_t\}$ and $\{\varepsilon_t\}$ are independent white noise series

– Then,

$$\nabla Y_t = \nabla M_t + \nabla e_t = \varepsilon_t + e_t - e_{t-1}$$

– Which would have the autocorrelation function of an MA(1) series

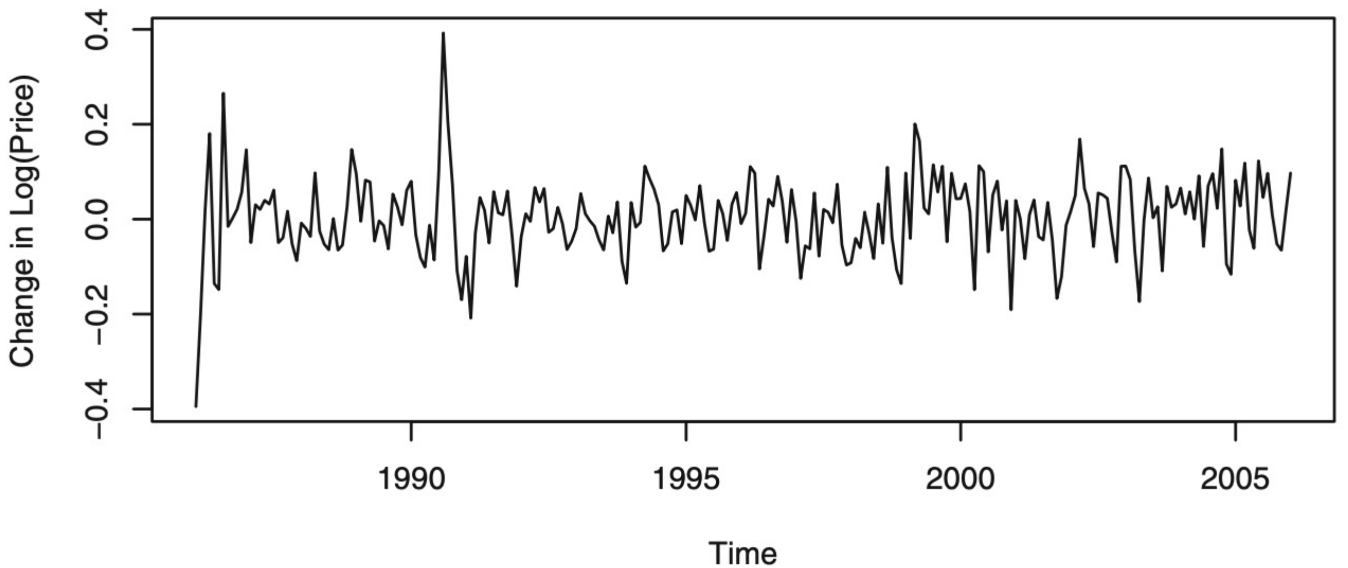
$$\rho_1 = -\{1/[2 + ((\sigma_\varepsilon^2)/(\sigma_e^2))]\}$$

– In either of these situations, we can study ∇Y_t as a stationary process

First difference

- Example: Oil price time series

Exhibit 5.4 The Difference Series of the Logs of the Oil Price Time



  - First differences of logarithms of the oil price time series
  - It looks much more stationary when compared with the original time series

Second difference

- We can easily extend this concept into second (or further) differencing. Assume that M_t is linear in time over three consecutive time points

  – Then, estimate M_t at the middle time point t by choosing $\beta_{0,t}$ and $\beta_{1,t}$ that achieve

$$\text{minimize} \sum_{j=-1}^{1} \left(Y_{t-j} - \left(\beta_{0,t} + j\beta_{1,t}\right)\right)^2$$

  – Which leads to

$$\widehat{M}_t = \frac{1}{3}(Y_{t+1} + Y_t + Y_{t-1})$$

Second difference

$Y_t = M_t + X_t$
M_t: approximately linear over every three consecutive time points
X_t: zero-mean stationary process

Estimate M_t by choosing β_0, β_1 which minimize $\sum_{j=-1}^{1}\left(Y_{t-j} - (\beta_{0,t} + \beta_{1,t})\right)^2$

- Then, the detrended series is

$$
\begin{aligned}
Y_t - \widehat{M}_t &= Y_t - \frac{1}{3}(Y_{t+1} + Y_t + Y_{t-1}) \\
&= -\frac{1}{3}(Y_{t+1} - 2Y_t + Y_{t-1}) \\
&= -\frac{1}{3}\nabla(\nabla Y_{t+1}) \\
&= -\frac{1}{3}\nabla^2 Y_{t+1}
\end{aligned}
$$

Note here that
$\nabla^2 Y_{t+1}$ is not $Y_{t+1} - Y_{t-1}$,
but $(Y_{t+1} - Y_t) - (Y_t - Y_{t-1})$

  – $\nabla^2 Y_{t+1}$ is the centered **second difference** of Y_t

- We can arrive at the similar situation when

$$Y_t = M_t + e_t \quad \text{where} \quad M_t = M_{t-1} + W_t \quad \text{and} \quad W_t = W_{t-1} + e_t$$

  – That is, M_t is a stochastic trend that its “rate of change (∇M_t)” is changing slowly over time

Chapter 5.2

ARIMA Models

ARIMA model

- A time series $\{Y_t\}$ is said to follow an **integrated auto-regressive moving average model** if the dth difference $W_t = \nabla^d Y_t$ is a stationary ARMA process

  – If $\{W_t\}$ follows an ARMA(p,q) model,
  we say that $\{Y_t\}$ is an **ARIMA(p,d,q)** process

  – (Fortunately, for practical purposes, we can usually take $d = 1$ or at most $d = 2$)

ARIMA(p,1,q)

- Consider an ARIMA(p,1,q) process. With $W_t = Y_t - Y_{t-1}$, we have

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

  – Or, in terms of the observed series,

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

  – Or, which we may rewrite as

$$Y_t = (1+\phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + \cdots + \left(\phi_p - \phi_{p-1}\right)Y_{t-p} - \phi_p Y_{t-p-1} \\ +e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

    - This is called the **difference equation form**

ARIMA(p,1,q)

$$Y_t = (1+\phi_1)Y_{t-1} + (\phi_2-\phi_1)Y_{t-2} + \cdots + \left(\phi_p - \phi_{p-1}\right)Y_{t-p} - \phi_p Y_{t-p-1}$$
$$+e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

- Notice that the difference equation form of an ARIMA(p,1,q) process appears to be an ARMA(p+1,q) process

- However, the characteristic polynomial satisfies

$$1-(1+\phi_1)x + (\phi_2-\phi_1)x^2 + \cdots + \left(\phi_p - \phi_{p-1}\right)x^p - \phi_p x^{p+1}$$
$$= (1-\phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p)(1-x)$$

  - This clearly shows the root at $x = 1$, which implies nonstationarity
  - However, the remaining roots are the roots of the characteristic polynomial of the stationary process ∇Y_t

IMA(1,1)

- The simple **IMA(1,1)** (or ARIMA(0,1,1)) model satisfactorily represents numerous time series, especially those arising in economics and business

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

  – To write Y_t explicitly as a function of present and past noise values (Here, we have to assume a couple of things. For more details, please see the textbook),

$$Y_t = e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1}$$

  – In contrast to stationary ARMA models, the weights on the white noise *do not die out* as we go into the past

    • Hence, Y_t can be regarded as an equally weighted accumulation of a large number of weighted white noise values

IMA(1,1)

- We can easily derive variances and correlations

$$Var(Y_t) = [1 + \theta^2 + (1 - \theta)^2(t + m)]\sigma_e^2$$

  – And

$$Corr(Y_t, Y_{t-k}) = \frac{[1 - \theta + \theta^2 + (1 - \theta)^2(t + m - k)]\sigma_e^2}{[Var(Y_t)Var(Y_{t-k})]^{1/2}}$$

$$\approx \sqrt{\frac{t+m-k}{t+m}} \approx 1 \quad \text{(for large } m \text{ and moderate } k\text{)}$$

  – As t increases, $Var(Y_t)$ increases and could be quite large
  – Correlation between Y_t and Y_{t-k} will be strongly positive for many lags $k = 1,2, \ldots$

IMA(2,2)

- An IMA(2,2) model can be written as

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

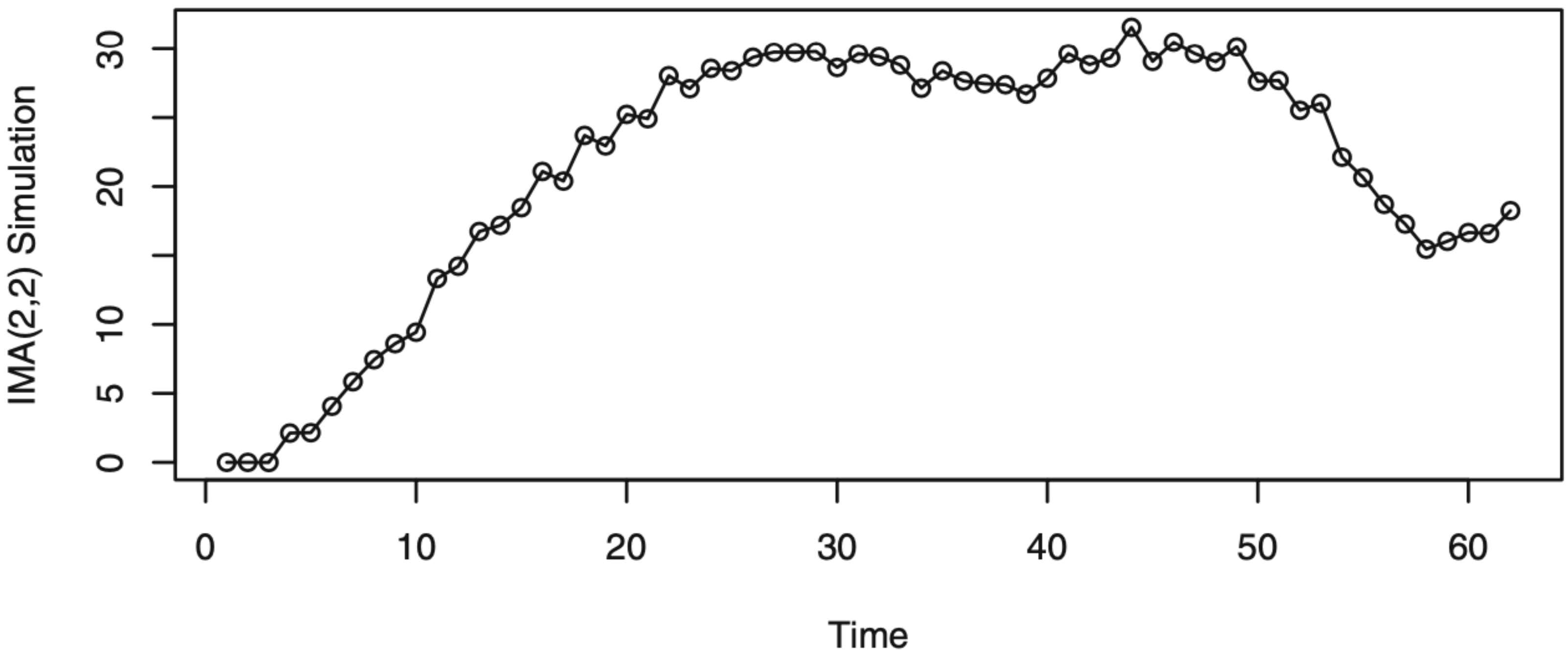
  - Or

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

  - We skip the calculations (because they are quite tedious), and we shall simply note that
    - Varaince of Y_t increases rapidly with t
    - Correlation between Y_t and Y_{t-k} is nearly 1 for all moderate k

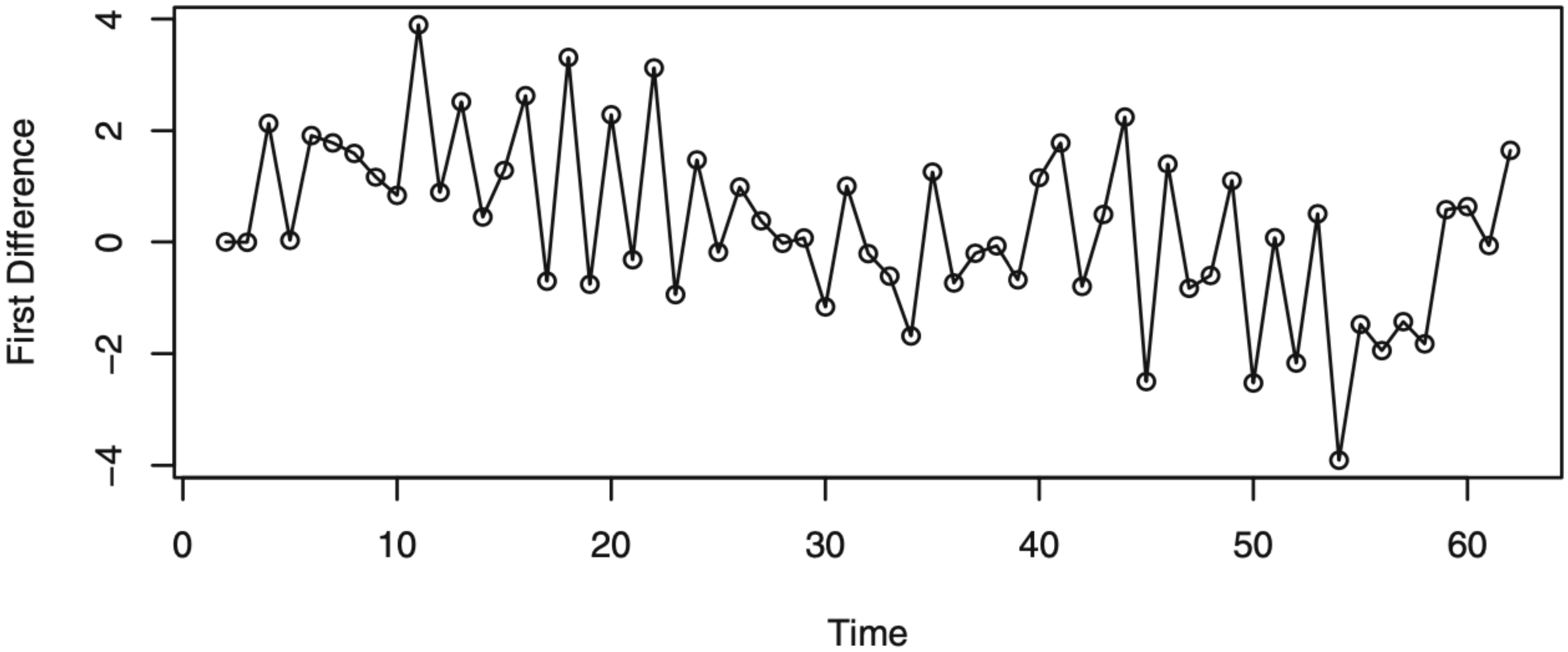
IMA(2,2)

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



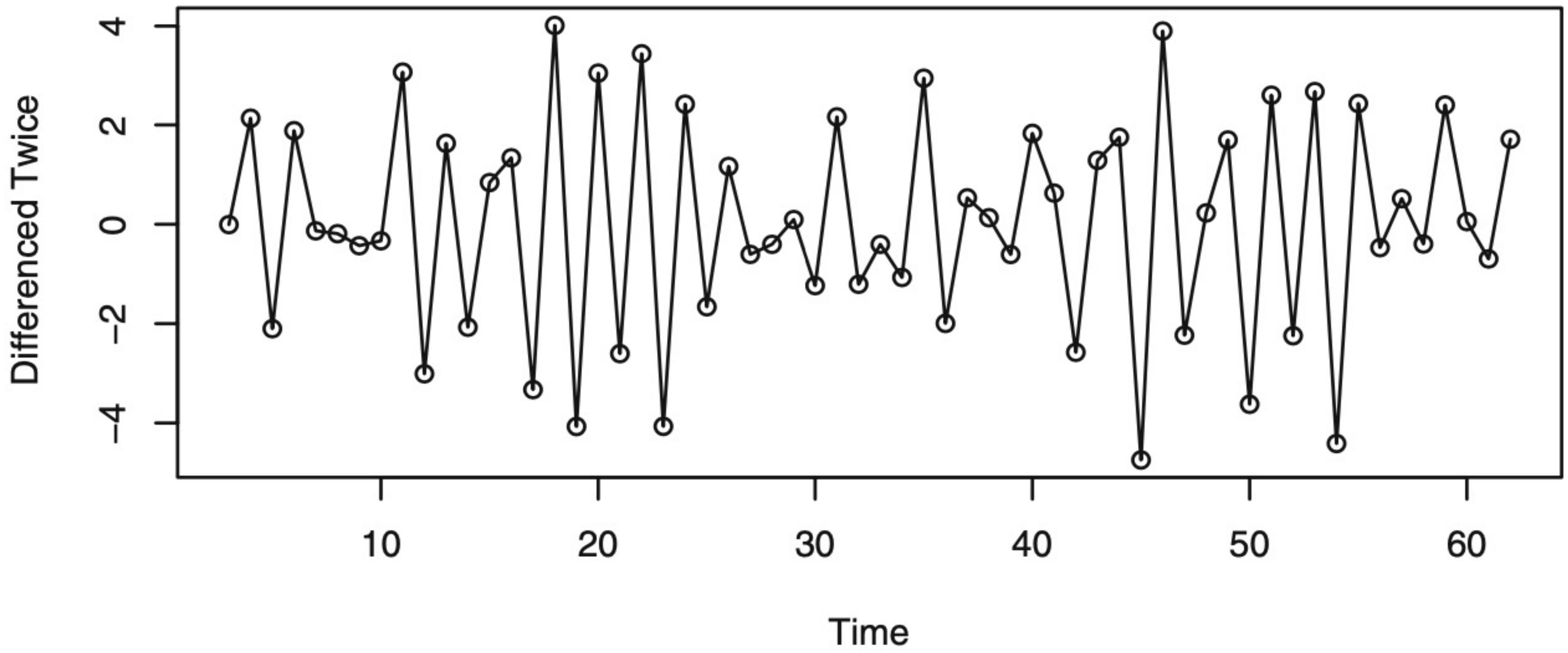
IMA(2,2)

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



IMA(2,2)

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



Chapter 5.3

Constant Terms in ARIMA Models

Constant terms in ARIMA models

- For an ARIMA(p,d,q) model, $\nabla^d Y_t = W_t$ is a stationary ARMA(p,q) process
  - Our standard assumption is that stationary models have a zero mean
  - What W_t has a nonzero mean μ?
    - First, we can just assume that

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

    - Or, we can introduce a constant term θ_0 into the model

$$W_t = \theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} \\ + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

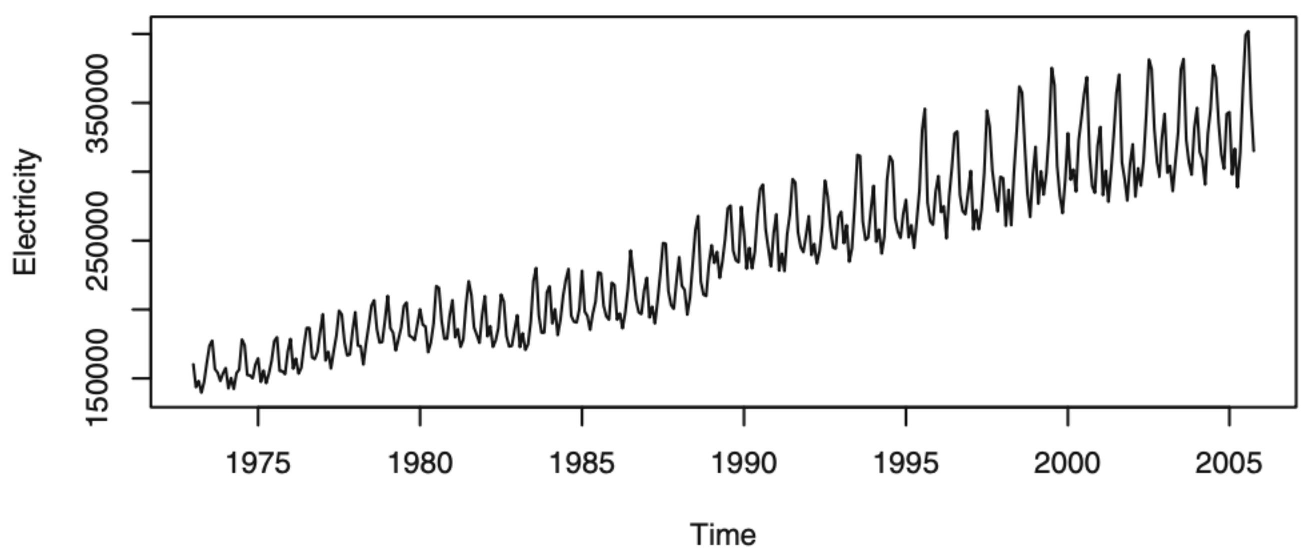
Chapter 5.4

Other Transformations

Logarithm

- We frequently encounter series where increased dispersion seems to be associated with higher level of the series
  - The higher the level of the series, the more variation there is around that level and conversely

Exhibit 5.8 U.S. Electricity Generated by Month



Logarithm

- We frequently encounter series where increased dispersion seems to be associated with higher level of the series
  - The higher the level of the series, the more variation there is around that level and conversely
  - Specifically, suppose that $Y_t > 0$ for all t and that

$$E(Y_t) = \mu_t \quad \text{and} \quad \sqrt{Var(Y_t)} = \mu_t \sigma$$

  - Then, consider the Taylor expansion of $\log(Y_t)$ around μ_t

$$\log(Y_t) \approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}$$

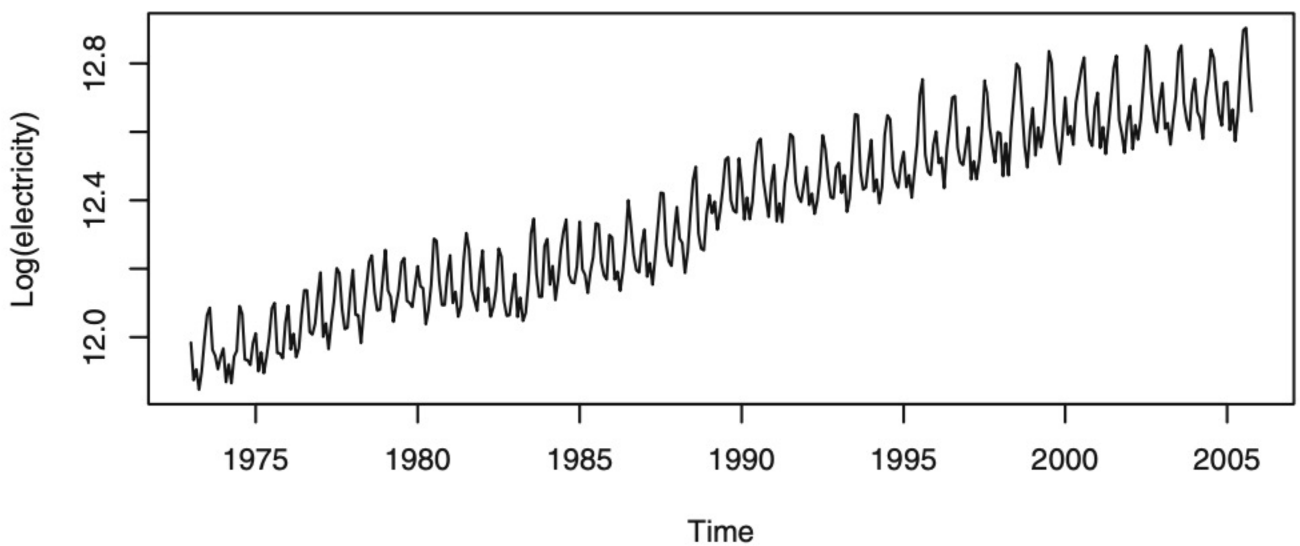
  - Taking the expected values and variances of both sides,

$$E[\log(Y_t)] \approx \log(\mu_t) \quad \text{and} \quad Var(\log Y_t) \approx \sigma^2$$

Logarithm

- We frequently encounter series where increased dispersion seems to be associated with higher level of the series
  - The higher the level of the series, the more variation there is around that level and conversely

Exhibit 5.9 Time Series Plot of Logarithms of Electricity Values



Percentage changes and logarithms

- Suppose Y_t tends to have relatively stable percentage changes from one time period to the next

$$Y_t = (1 + X_t)Y_{t-1}$$

  – Where $100X_t$ is the percentage change (possibly negative) from Y_{t-1} to Y_t (e.g., $X_t = 0.1$ means 10% increase)

$$\log(Y_t) - \log(Y_{t-1}) = \log\left(\frac{Y_t}{Y_{t-1}}\right) = \log(1 + X_t)$$

  – If X_t is restricted to, say, $|X_t| < 0.2$ (i.e. $\pm$20%),

$$\log(1 + X_t) \approx X_t$$

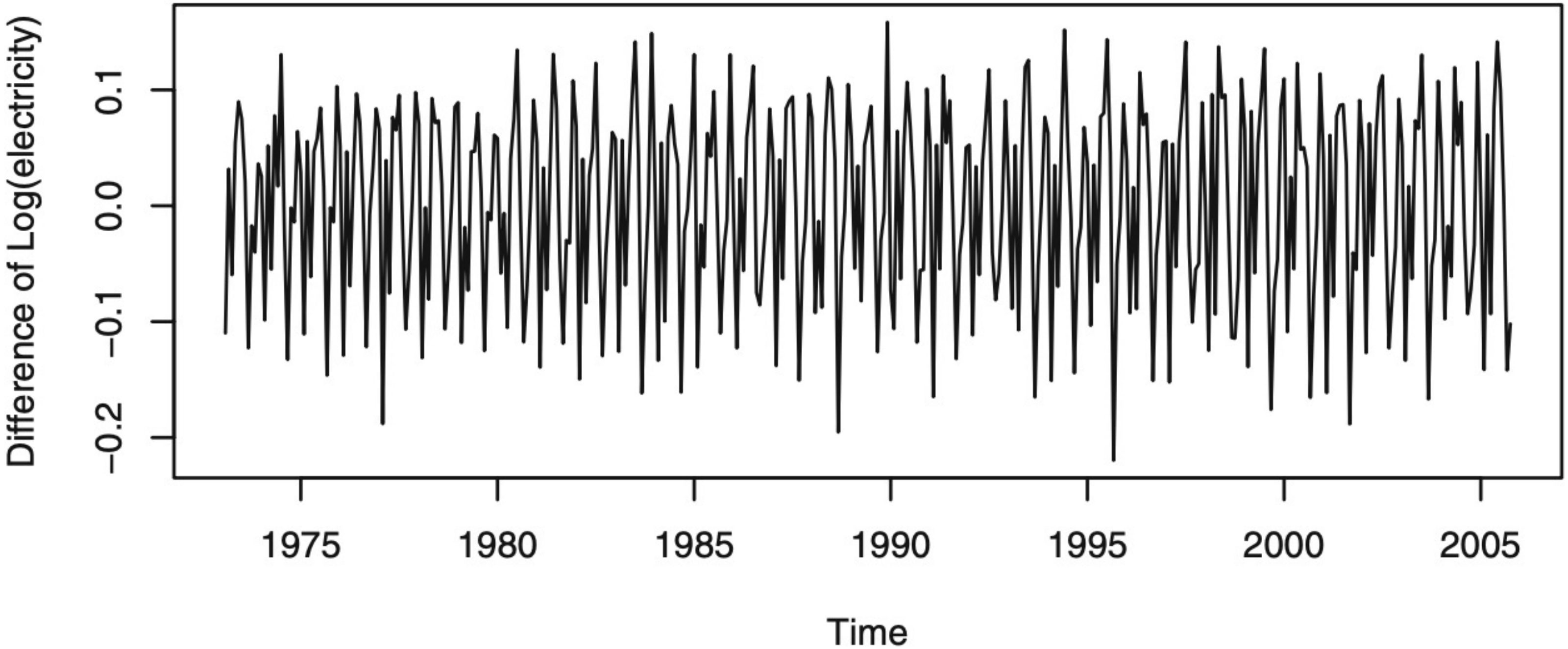
  – That is, the first differences of logarithms are close to percentage changes

$$\nabla[\log(Y_t)] \approx X_t$$

These are called 'log returns' in finance

Percentage changes and logarithms

Exhibit 5.10 Difference of Logarithms for Electricity Time Series



Power transformations

- A more general **power transformation** was introduced by Box and Cox (1964)
  - For a given value of the parameter λ,

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