DS-GA 3001.001
Probabilistic time series analysis
Lecture 3
AR(I)MA

Instructor: Cristina Savin

The backshift operator

Definition 2.4 *We define the* **backshift operator** *by*

$$Bx_t = x_{t-1}$$

and extend it to powers $B^2x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$, *and so on. Thus,*

$$B^k x_t = x_{t-k}.$$

Inverse (forward-shift operator)

$$x_t = B^{-1}Bx_t = B^{-1}x_{t-1}$$

Finite differences:

$$\nabla x_t = (1 - B)x_t$$

$$\nabla^d = (1 - B)^d$$

Compact way to write complex models

Quick way to check for stationarity, causality

Helps solve issues with degeneracies and overparametrization

General framework for computing ACF

Go back to AR(1), rewrite using backshift operator

Rewrite equation

$$
\begin{aligned}
X_t - \lambda X_{t-1} &= W_t \\
(1 - \lambda B) X_t &= W_t \\
P(B) X_t &= W_t
\end{aligned}
$$

Using B powers:

$$X_t = \sum_{k=0} \lambda^k W_{t-k} = \underbrace{\boxed{\sum_{k=0}^{\infty} \lambda^k B^k}}_{Q(B)} W_t$$

$$X_t = \lambda X_{t-1} + W_t$$

What happens when $|\lambda| > 1$?

$Q(B)W_t = \sum_{k\geq 0} \lambda^k B^k W_t$ does not converge

But we can rewrite everything (essentially flipping time axis)

$$\frac{1}{\lambda}X_t = \frac{\lambda}{\lambda}X_{t-1} + \frac{1}{\lambda}W_t$$

$$X_{t-1} = \lambda^{-1}X_t - \lambda^{-1}W_t$$

Anti-causal : future determines the past

$$X_t = -\sum_{k=1}^{\infty} \lambda^{-k} W_{t+k}$$

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \text{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have
$$\begin{aligned} X_t &= P(B)^{-1} W_t \\ &= Q(B) W_t \end{aligned}$$

Operators **P** and **Q** behave like regular polynomials

$$\frac{1}{1 - \lambda z} = \sum_{k \geq 0} \lambda^k z^k \ , \ |\lambda| < 1, |z| \leq 1$$

Stationarity and causality

Theorem

1. *The equation $P(B)X_t = W_t$ has a unique stationary solution if and only if*

$$P(z) = 0 \Rightarrow |z| \neq 1 \ .$$

   We call this unique solution an AR(p) process.

2. *Moreover, this process is causal if and only if*

$$P(z) = 0 \Rightarrow |z| > 1 \ .$$

Roots of polynomial determine properties of the stochastic process

Why is this useful? Degeneracy in parametrization

$$X_t = W_t + \lambda W_{t-1}$$

$$R_X(t, t+h) = \begin{cases} \sigma^2\left(1+\lambda^2\right), & h = 0 \\ \sigma^2\lambda, & |h| = 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\sigma^2 = 1$$

$$\lambda = 5$$

$$\sigma^2 = 25$$

$$\lambda = \frac{1}{5}$$

Revisiting MA(1)

$$X_t = W_t + \theta W_{t-1} = (1 + \theta B) W_t = P(B) W_t$$

$|\theta| < 1,$

$$\begin{aligned} P(B)^{-1} X_t &= W_t \\ \frac{1}{1+\theta B} X_t &= W_t \\ \left(1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \ldots\right) X_t &= W_t \\ \sum_{k \geq 0} (-\theta)^k X_{t-k} &= W_t \ , \end{aligned}$$

essentially, we have inverted the roles of X and W

DEF: Invertible Process

A linear process $\{X_t\}$ is **invertible** if there exist $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$ with $\sum_k |\psi_k| < \infty$ and

$$\psi(B)X_t = W_t \ .$$

AR(1)

$X_t - \lambda X_{t-1} = (1 - \lambda B)X_t = W_t$

Causal (wrt $\{W_t\}$) iff $|\lambda| < 1$.

Always invertible (wrt $\{W_t\}$).

MA(1)

$X_t = W_t + \theta W_{t-1} = (1 + \theta B)W_t$

Always causal (wrt $\{W_t\}$).

Invertible (wrt $\{W_t\}$) iff $|\theta| < 1$.

Increasing complexity: AR(p)

An AR(p) process $\{X_t\}$ is a stationary process satisfying

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$P(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \ldots \lambda_p B^p$$

Constraints on polynomial P(B)

$|z_k^*| \neq 1$ for all (complex) roots of P(B)

Polynomials refresher

A polynomial of order n has n complex roots

If coeff. are real valued- pairs of conjugate roots

Increasing complexity: MA(q)

The moving average model of order q, or MA(q), is defined as

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \dots \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\theta_q \neq 0$$

$$X_t = \theta(B) W_t$$

$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$\lambda_p \neq 0 \ \ \theta_q \neq 0$

The **autoregressive operator** *is defined to be*

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p .$$

The **moving average operator** *is*

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q .$$

$$\boxed{\phi(B) x_t = \theta(B) w_t .}$$

***NO COMMON ROOTS**

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$\phi(B)x_t = \theta(B)w_t .$$

Where $\{W_t\}$ is white noise

*no **common** roots

Special cases

AR(p) $=$ ARMA(p, 0)

MA(q) $=$ ARMA(0,q)

Has p+q parameters

For any stationary process with autocovariance R and any $k > 0$, there is an ARMA process $\{X_t\}$ such that

$$R_X(h) = R(h) \ , h \leq k \ .$$

The wonderful world of ARMA polynomials

$$P(B)X_t = \theta(B)W_t$$

Where P(B) has degree p and
Q(B) has degree q

We can think an ARMA model as concatenating two models:

$$Y_t = \theta(B)W_t \ , \ \text{and} \ \ P(B)X_t = Y_t \ .$$

Theorem

- *If P and θ have no common factors, a stationary solution to $P(B)X_t = \theta(B)W_t$ exists iff the roots of $P(z)$ avoid the unit circle: $P(z) = 0 \Rightarrow |z| \neq 1$. This is called an ARMA(p,q) process.*
- *This process is* **causal** *iff the roots of $P(z)$ are* **outside** *the unit circle: $P(z) = 0 \Rightarrow |z| > 1$.*
- *This process is* **invertible** *iff the roots of $\theta(B)$ are* **outside** *the unit circle: $\theta(z) = 0 \Rightarrow |z| > 1$.*

Example: minimal models

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

$$\theta(B) = (1 + B + .25B^2) = (1 + .5B)^2$$

$$\phi(B) = 1 - .4B - .45B^2 = (1 + .5B)(1 - .9B)$$

Simplified, this is actually ARMA(1,1)

$$x_t = .9x_{t-1} + .5w_{t-1} + w_t$$

How do we use B to compute the ACF?

Self-consistency of CCF/ACF: difference equations

AR(1): $x_t = \phi x_{t-1} + w_t$

$$\rho(h) = \phi\, \rho(h-1), \quad h = 1, 2, \ldots .$$

Homogeneous difference equation of order 1:

$$u_n - \alpha u_{n-1} = 0, \quad \alpha \neq 0, \quad n = 1, 2, \ldots .$$

Self-consistency of CCF/ACF: difference equations

AR(2): $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$

Homogeneous difference equation of order 2:

$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0, \quad \alpha_2 \neq 0, \quad n = 2, 3, \ldots$$

$$\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2,$$

$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n},$$ **Or** $$u_n = z_0^{-n}(c_1 + c_2 n).$$

In general:

Homogeneous difference equation of order p:

$$u_n - \alpha_1 u_{n-1} - \cdots - \alpha_p u_{n-p} = 0, \quad \alpha_p \neq 0, \quad n = p, p+1, \ldots\ .$$

$$\begin{aligned} u_n &= z_1^{-n} \times (\text{a polynomial in } n \text{ of degree } m_1 - 1) \\ &+\ z_2^{-n} \times (\text{a polynomial in } n \text{ of degree } m_2 - 1), \\ &\quad +\ldots \end{aligned}$$

How about ARMA models?

$$\gamma(h) - \phi_1\gamma(h-1) - \cdots - \phi_p\gamma(h-p) = 0, \quad h \geq \max(p, q+1),$$

with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j\gamma(h-j) = \sigma_w^2 \sum_{j=h}^{q} \theta_j\psi_{j-h}, \quad 0 \leq h < \max(p, q+1).$$

See tsa4.pdf Section 3.3

Example 3.14 The ACF of an ARMA(1, 1)

$$\gamma(h) - \phi\gamma(h-1) = 0, \quad h = 2, 3, \ldots,$$

$$\gamma(h) = c\,\phi^h, \quad h = 1, 2, \ldots .$$

$$\rho(h) = \frac{(1+\theta\phi)(\phi+\theta)}{1+2\theta\phi+\theta^2}\,\phi^{h-1}, \quad h \geq 1.$$

$$\gamma(0) = \phi\gamma(1) + \sigma_w^2[1+\theta\phi+\theta^2] \quad \text{and} \quad \gamma(1) = \phi\gamma(0) + \sigma_w^2\theta.$$

General linear model parametrization

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots)(\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots) = (1 + \theta_1 z + \theta_2 z^2 + \cdots).$$

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 - \phi_1\psi_0 &= \theta_1 \\
\psi_2 - \phi_1\psi_1 - \phi_2\psi_0 &= \theta_2 \\
\psi_3 - \phi_1\psi_2 - \phi_2\psi_1 - \phi_3\psi_0 &= \theta_3 \\
&\vdots
\end{aligned}$$

Yet another h.d.e:

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q+1),$$

with initial conditions

$$\psi_j - \sum_{k=1}^{j} \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j < \max(p, q+1).$$

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

In general, it may take several differentiations to get there (d-th order polynomial dependence on time)

A process is **ARIMA** (p,**d**,q) if **d**-th difference

$$\nabla^d x_t = (1 - B)^d x_t$$ is ARMA (p,q)

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t$$

These are tricky to use, motivate state space models

Revisiting inference

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \;, |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

minimum->
derivatives zero
(check at home, tsa4 theorem B3)

$$\alpha = \mu(1 - \rho_X(h)) \;,\; \beta = \rho_X(h)$$

$$MSE = \sigma^2(1 - \rho_X(h)^2)$$

Optimal **linear** predictor

$$f(x_t) = \mu + \rho_X(h)(x_t - \mu)$$

The optimal predictor
if stationary gaussian

Best Linear Predictor

$$x_{n+1}^n = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \cdots + \phi_{nn}x_1,$$

$$\phi_n = \Gamma_n^{-1}\gamma_n.$$

$$x_{n+1}^n = \phi_n' x,$$

$$P_{n+1}^n = \mathrm{E}(x_{n+1} - x_{n+1}^n)^2 = \gamma(0) - \gamma_n'\Gamma_n^{-1}\gamma_n.$$

Property 3.4 The Durbin–Levinson Algorithm

Equations (3.64) and (3.66) can be solved iteratively as follows:

$$\phi_{00} = 0, \quad P_1^0 = \gamma(0).$$

For $n \geq 1$,

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1}\phi_{n-1,k}\ \rho(n-k)}{1 - \sum_{k=1}^{n-1}\phi_{n-1,k}\ \rho(k)}, \quad P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2),$$

where, for $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, \ldots, n-1.$$

Property 3.5 Iterative Solution for the PACF

The PACF of a stationary process x_t, can be obtained iteratively via (3.69) as ϕ_{nn}, for $n = 1, 2, \ldots$.

Property 3.6 The Innovations Algorithm

The one-step-ahead predictors, x_{t+1}^t, and their mean-squared errors, P_{t+1}^t, can be calculated iteratively as

$$x_1^0 = 0, \quad P_1^0 = \gamma(0)$$

$$x_{t+1}^t = \sum_{j=1}^{t} \theta_{tj}(x_{t+1-j} - x_{t+1-j}^{t-j}), \quad t = 1, 2, \ldots \tag{3.77}$$

$$P_{t+1}^t = \gamma(0) - \sum_{j=0}^{t-1} \theta_{t,t-j}^2 P_{j+1}^j \quad t = 1, 2, \ldots, \tag{3.78}$$

where, for $j = 0, 1, \ldots, t-1$,

$$\theta_{t,t-j} = \left(\gamma(t-j) - \sum_{k=0}^{j-1} \theta_{j,j-k}\theta_{t,t-k}P_{k+1}^k\right) \Big/ P_{j+1}^j. \tag{3.79}$$

Given data $x_1, \ldots, x_n$, the innovations algorithm can be calculated successively for $t = 1$, then $t = 2$ and so on, in which case the calculation of x_{n+1}^n and P_{n+1}^n is made at the final step $t = n$. The m-step-ahead predictor and its mean-square error based on the innovations algorithm (Problem 3.46) are given by

$$x_{n+m}^n = \sum_{j=m}^{n+m-1} \theta_{n+m-1,j}(x_{n+m-j} - x_{n+m-j}^{n+m-j-1}), \tag{3.80}$$

$$P_{n+m}^n = \gamma(0) - \sum_{j=m}^{n+m-1} \theta_{n+m-1,j}^2 P_{n+m-j}^{n+m-j-1}, \tag{3.81}$$

where the $\theta_{n+m-1,j}$ are obtained by continued iteration of (3.79).

How do we estimate the model parameters?

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Assume: model known, and zero mean (preprocessing)

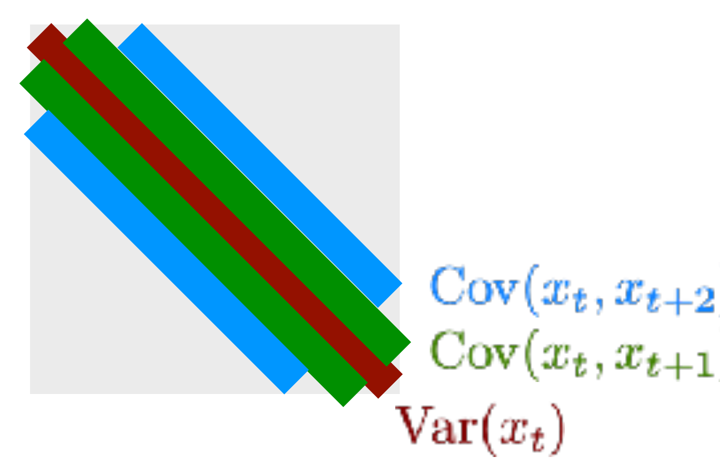
Gaussian stats $P(B)X_t = \theta(B)W_t$ Where $\{W_t\}$ is iid gaussian noise 0 mean, σ^2 variance

Find parameters $\lambda_i,\ \Theta_j\ \sigma^2$ that maximize likelihood

$$\mathcal{L}(\lambda, \theta, \sigma^2) = f_{\lambda,\theta,\sigma^2}(x_1, \ldots, x_n)$$

Jointly gaussian

$$\mathcal{L}(\lambda, \theta, \sigma^2) = (2\pi)^{-n/2}|\Gamma_n|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^T\Gamma_n^{-1}\mathbf{x}\right)$$



Where we've collated the data in vector $\mathbf{x} = (x_1, \ldots, x_n)$

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

Cons: unpleasant optimization, need good initialization

We need to find a cheap initial guess **Yule Walker equations**

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t,$$

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p), \quad h = 1, 2, \ldots, p,$$
$$\sigma_w^2 = \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p).$$

Solve for parameters:

$$\Gamma_p \phi = \gamma_p, \quad \sigma_w^2 = \gamma(0) - \phi' \gamma_p,$$

Replace with empirical estimates

$$\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p, \quad \hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\gamma}_p' \hat{\Gamma}_p^{-1} \hat{\gamma}_p.$$

Nice asymptotic convergence properties, see book for details

Beyond ARIMA

Seasonal ARIMA:

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t, \tag{3.158}$$

where the operators

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps} \tag{3.159}$$

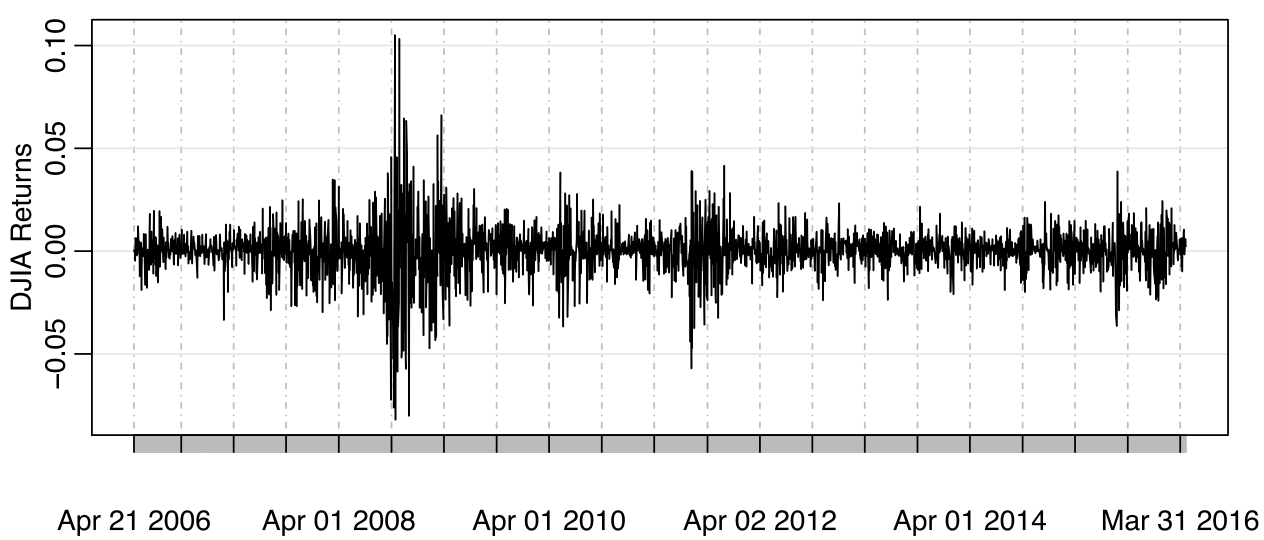
and

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs} \tag{3.160}$$

are the **seasonal autoregressive operator** and the **seasonal moving average operator** of orders P and Q, respectively, with seasonal period s.

Modeling changes in variance (ARCH, GARCH)

autoregressive conditionally heteroscedastic



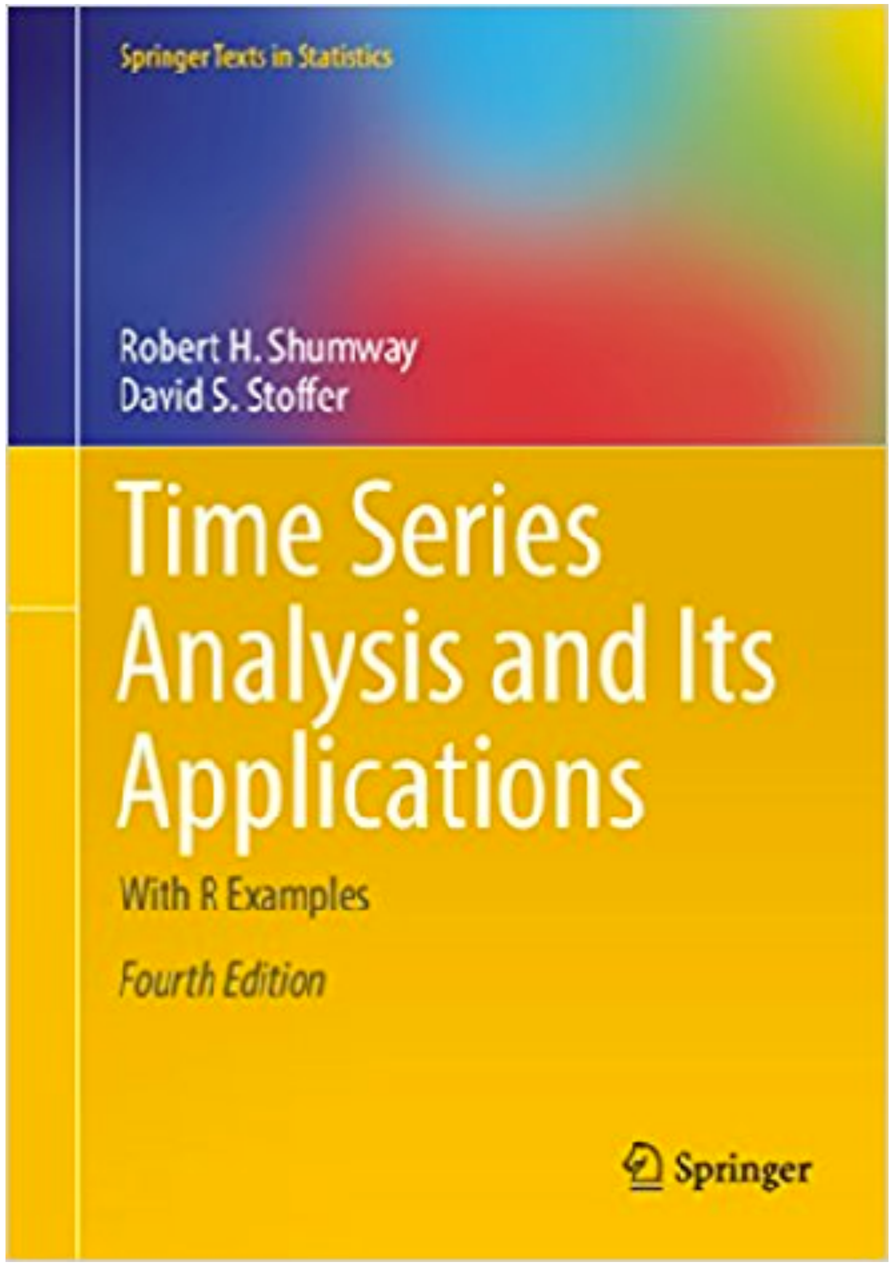
$$\nabla \log(x_t) \approx r_t.$$

$$r_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2,$$

DJIA: *Dow Jones Industrial Average*

Chapter 1 & 3



Homework 1:
due Sept.29