MSDS 596 Regression & Time Series

Lecture 11 Time Series Model Specification

Department of Statistics
Rutgers University

Nov 12, 2020

Schedule

Week	Date	Topic
1	9/3	Intro to linear regression (JF1,2)
2	9/10	Estimation (JF2)
3	9/17	Inference I (JF3)
4	9/24	Inference II (JF3)
5	10/1	Inference and prediction (JF4,5)
6	10/8	Explanation; model diagnostics (JF6-8)
7	10/15	Transformation and model selection (JF9-10)
8	10/22	Shrinkage methods (JF11)
9	10/29	Time series exploratory analysis (CC2,3)
10	11/5	Linear time series: ARIMA models (CC4-5)
11	11/12	Model specification and estimation (CC6,7)
12	11/19	Diagnostics and forecasting (CC8,9)
	11/26	(no class)
13	12/3	Seasonal models (CC10)
14	week of 12/7	Project & final evaluation

Recall: General Linear Process - ARMA Processes

Let $\{Y_t\}$ be an observed time series and $\{e_t\}$ an unobserved white noise.

- $\{Y_t\}$ is said to be a **general linear process** if it can be represented as

$$y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots,$$

  with $\sum_{i=1}^{\infty} \psi_i^2 < \infty$ and $\psi_0 = 1$.
- $\{Y_t\}$ is said to be an **ARMA(p, q)** process, if for every t

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

- Characteristic polynomials of the AR(p) and MA(q) components

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p,$$
$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q.$$

- The ARMA(p, q) process is stationary if its AR(p) component is stationary, i.e. if all roots to $\phi(z) = 0$ are larger than 1 in modulus.
- The ARMA(p, q) process is invertible if its MA(q) component is invertible, i.e. if all roots to $\theta(z) = 0$ are larger than 1 in modulus.

ARMA(p, q)

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

- The GLP representation of ARMA(p, q) is

$$
\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= -\theta_1 + \phi_1 \\
\psi_2 &= -\theta_2 + \phi_2 + \phi_1 \psi_1 \\
&\vdots \\
\psi_k &= -\theta_k + \phi_p \psi_{k-p} + \phi_{p-1} \psi_{k-p+1} + \cdots + \phi_1 \psi_{k-1}
\end{aligned}
$$

  assuming $\psi_k = 0$ for all $k < 0$, and $\theta_k = 0$ for all $k > q$.
- The ACF of ARMA(p, q) satisfies for all $k > q$,

$$
\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}.
$$

  For $k \leq q$, ρ_k also involves the θ terms.

Nonstationary time series

- Example: a time series $\{Y_t\}$ is called a random walk if it satisfies:

$$Y_t = Y_{t-1} + e_t, \qquad \nabla Y_t = Y_t - Y_{t-1} = e_t$$

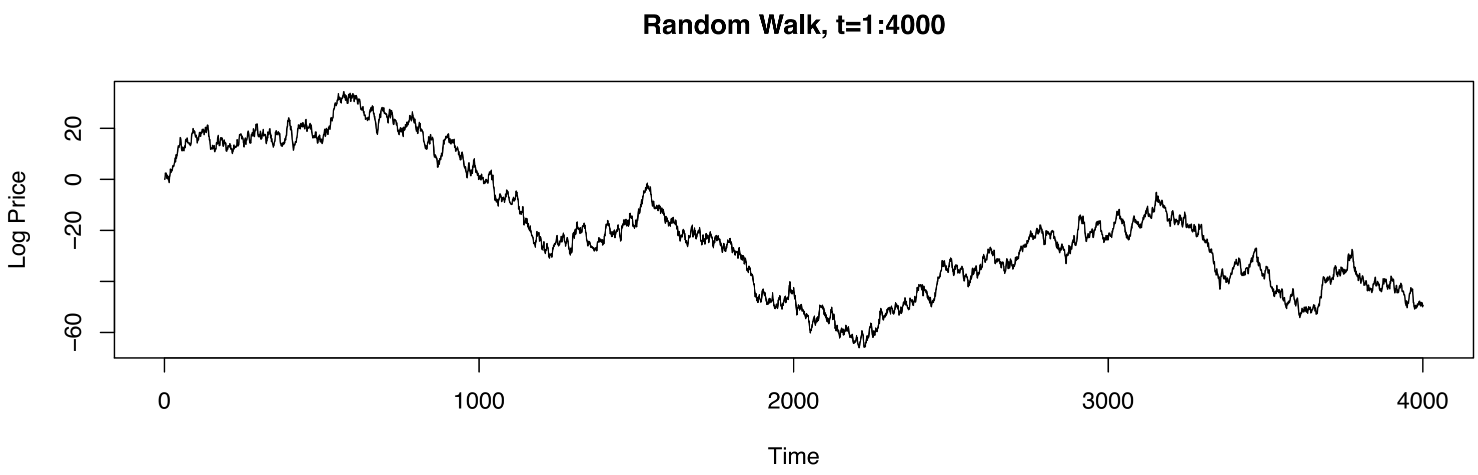
  where e_t are i.i.d. with mean zero and variance σ^2.
- It is an AR(1) model with coefficient $\phi_1 = 1$ and $\phi_0 = 0$. AR(1): $\rho_k = \phi^k$
  - Nonstationary: the variance diverges as t increases.
  - Strong memory: the sample ACF approaches 1 for any finite lag.
  - *Unpredictable: l-step ahead forecast is $\hat{r}_h(l) = Y_h$, and $\mathrm{Var}(e_h(l)) = l\sigma^2$.
- A random walk with drift takes the form $Y_t = \mu + Y_{t-1} + e_t$.
  - Same properties as a random walk;
  - In addition, it has a time trend with slope μ:

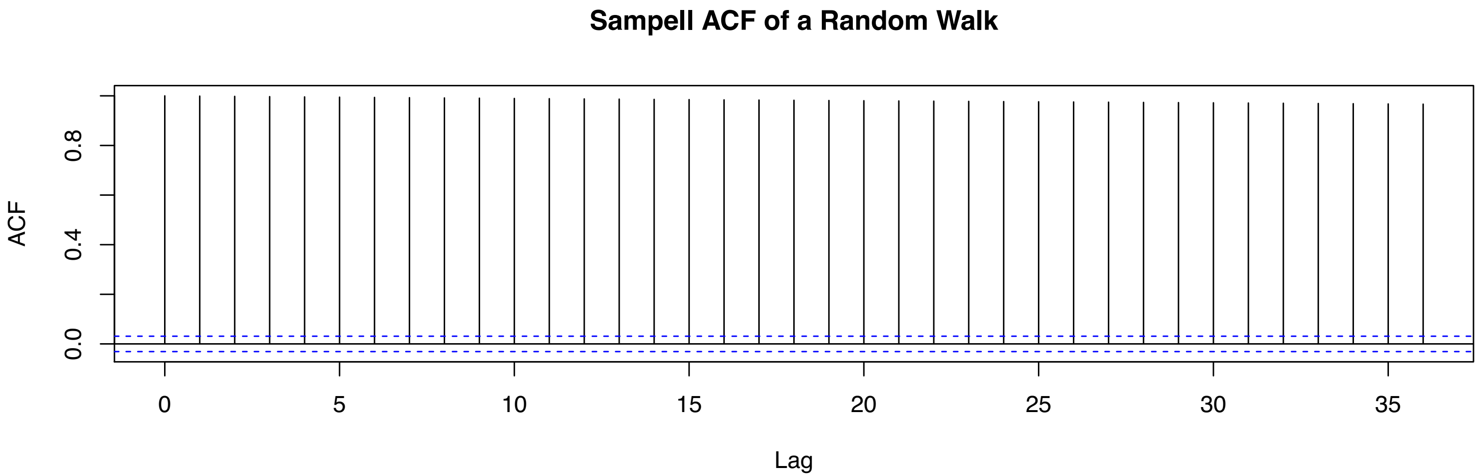
$$\nabla Y_t = Y_t - Y_{t-1} = \mu + e_t$$

$$Y_t = \mu t + Y_0 + e_1 + e_2 + \cdots + e_t.$$

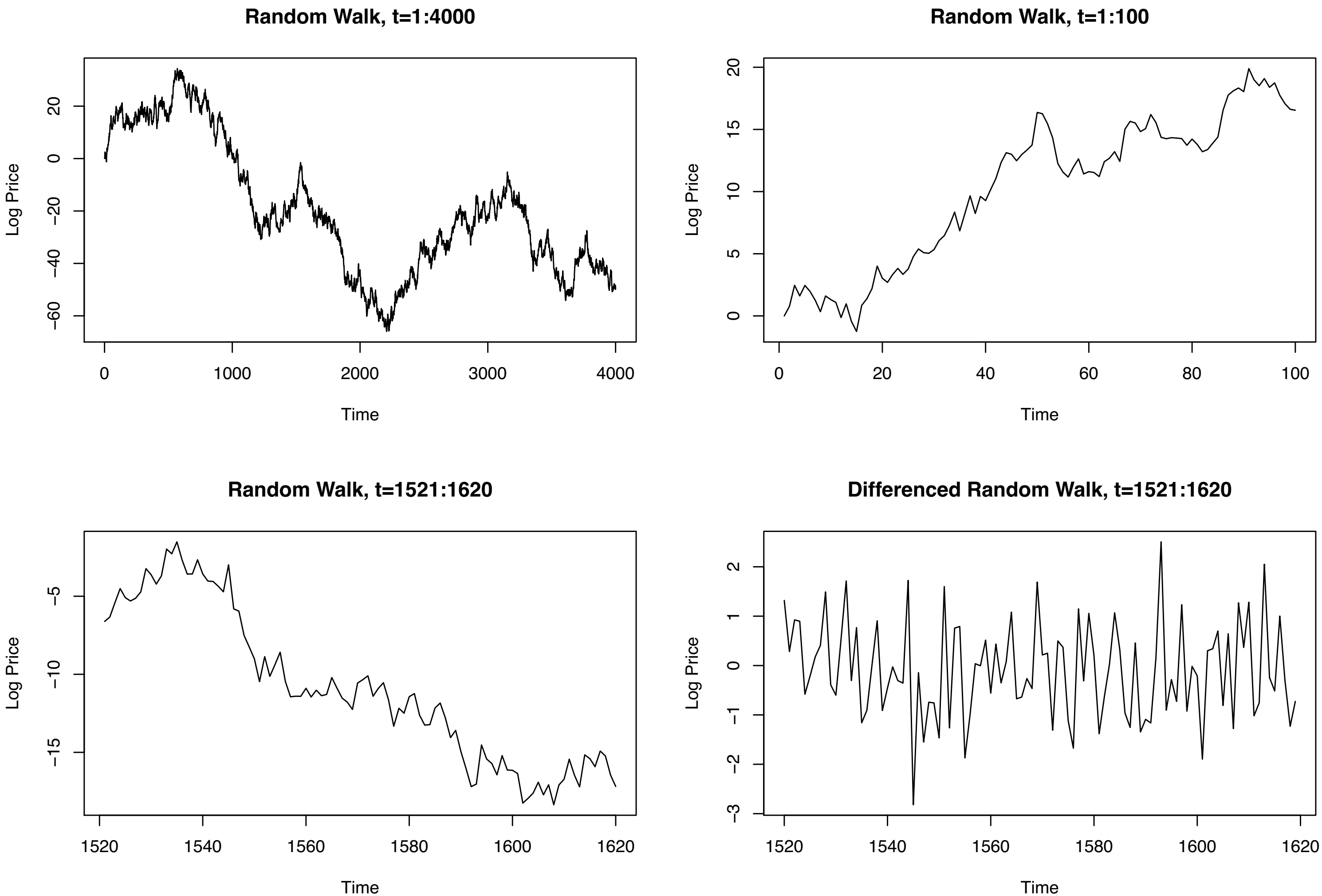
- Differencing: $\nabla Y_t := Y_t - Y_{t-1}$ leads to a white noise (with nonzero mean, if there is a drift).

Example: ACF does not decay

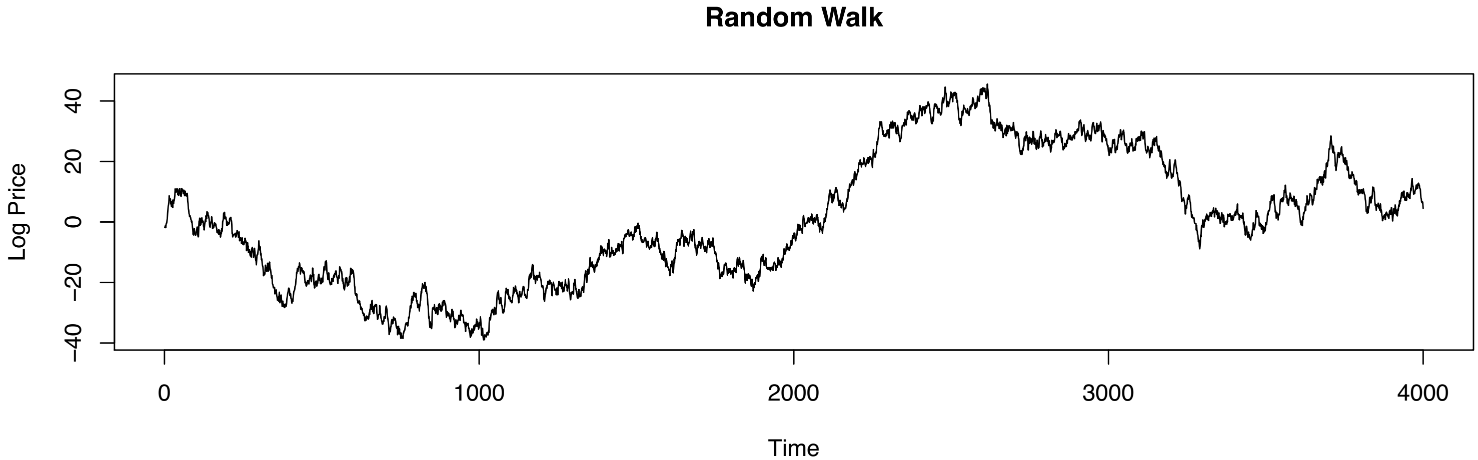


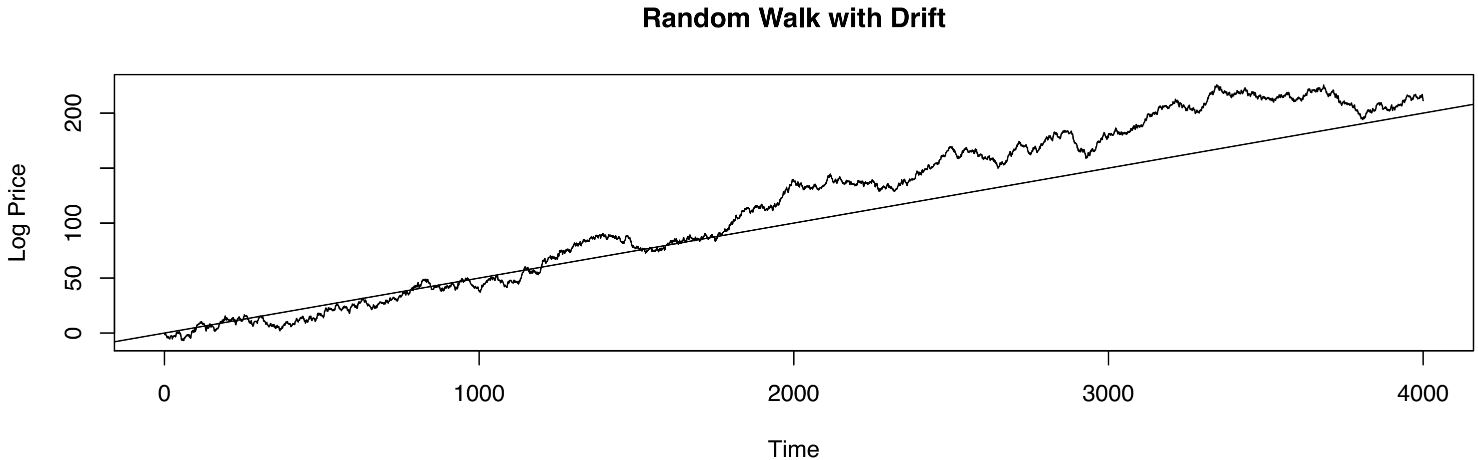


Example: Trend over a Short Time Period



Example: Random Walk with a Drift





The Backshift Operator

The **backshift operator** B operates on the time index of a series and shifts time back one time unit to form a new series. That is,

$$Y_{t-1} = BY_t.$$

- B is linear: $B(aY_t + bX_t + c) = aBY_t + bBX_t + c$; $= aY_{t-1} + bX_{t-1} + c$
- For any positive integer m, $B^m Y_t = Y_{t-m}$. $= B \cdot B \cdots BY_t$
- A general AR(p) process can be written as

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

  where $\phi(B)$ is the AR characteristic polynomial evaluated at B;
- A general MA(q) process can be written as

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

  where $\theta(B)$ is the MA characteristic polynomial evaluated at B;
- Combining the two, a general ARMA(p, q) process can be written as

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

The Differencing Operator

The **differencing operator** $\nabla = 1 - B$. Therefore,

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

- The **second difference**

$$\nabla^2 Y_t = (1 - B)^2 Y_t = (1 - B)(Y_t - Y_{t-1}) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2});$$

- The d**-th difference** $\nabla^d Y_t = (1 - B)^d Y_t$;

! To be distinguished from the **seasonal difference** of period s:

$$\nabla_s Y_t = (1 - B^s) Y_t = Y_t - Y_{t-s}.$$

ARIMA Models

- A time series $\{Y_t\}$ is said to be an **autoregressive integrated moving average process of order** $(p, 1, q)$, denoted ARIMA(p, 1, q), if the differenced series $W_t = \nabla Y_t = Y_t - Y_{t-1}$ follows a ARMA(p, q) model.
- In general, a time series $\{Y_t\}$ is said to be an **autoregressive integrated moving average process of order** (p, d, q), denoted ARIMA(p, d, q), if the d^{th} difference $\nabla^d Y_t$ follows a ARMA(p, q) model.

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

Examples of ARIMA Models

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

- Random walk, i.e. ARIMA(0, 1, 0), or the I(1) process: $e_t = \nabla Y_t$.
- IMA(1, 1) = ARIMA(0, 1, 1): $W_t = \nabla Y_t = e_t - \theta e_{t-1}$;
- ARI(1, 1) = ARIMA(1, 1, 0): $\nabla Y_t = \phi \nabla Y_{t-1} + e_t$, for $|\phi| < 1$;
- IMA(2, 2): $W_t = \nabla^2 Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$.

Constant term in ARIMA models

Let $\{Y_t\}$ be an ARIMA(p, d, q) process, that is, $W_t = \nabla^d Y_t$ follows

$$\begin{aligned} W_t &= \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \end{aligned}$$

which assumes $E(W_t) = 0$.
How to express $E(W_t) = \mu \neq 0$? Two ways to rewrite the model:

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

and

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

The fact that the two ways are equivalent means

$$\mu = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \Leftrightarrow \theta_0 = \mu\left(1 - \phi_1 - \phi_2 - \cdots - \phi_p\right)$$

"intercept" in R model fit

Constant term in ARIMA models: examples

What is $E(Y_t)$ for each of the following constant-added models?

- AR(p): $Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t$;
- MA(q): $Y_t = \theta_0 + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$;
- IMA(1, 1): $Y_t = Y_{t-1} + \theta_0 + e_t - \theta e_{t-1}$, implying that

$$Y_t = e_t + (1-\theta)e_{t-1} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1} + (t+m+1)\theta_0,$$

that is, there's an additional deterministic trend linear in t.

Alternative representation of ARIMA(p,d,q) with constant term:

$$Y_t = Y'_t + \mu_t,$$

where $\{Y'_t\}$ follows ARIMA(p,d,q) with $E(\nabla^d Y'_t) = 0$, and μ_t is a polynomial in t of degree d.

E.g. random walk &
random walk w/ drift

Transformations

- Nonstationarity due to non-constant mean: differencing;
- Nonstationarity due to non-constant variance:
  - Suppose $E(Y_t) = \mu_t$, and $\sqrt{Var(Y_t)} = \mu_t \sigma$.
  - Then, $E(\log(Y_t)) = \log \mu_t$, and $Var(\log(Y_t)) \approx \sigma^2$ (by Taylor expansion).
- Percentage changes and logarithms:
  - Suppose Y_t has stable percentage change over time: $Y_t = (1 + X_t) Y_{t-1}$, where $|X_t|$ is small ($< 20\%$). Then,

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

    may be modeled as a stationary process.
- Power and Box-Cox transformations for time series.

ARIMA model fitting: general strategy

1. **Specification**: decide on reasonable and tentative values for (p, d, q);
2. **Estimation***: estimate parameters in the most efficient way;
3. **Diagnostics**: look critically at the fitted model just obtained to check its adequacy, such as we did for regression and deterministic trend time series models.

Parameter estimation for AR processes

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t.$$

Recall: ACF of AR(p) can be derived via the Yule-Walker equations:

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \cdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \cdots \\ \gamma_p \end{pmatrix},$$

and

$$\sigma^2 = \gamma_0 - \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_p \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \cdots \\ \gamma_p \end{pmatrix}.$$

Parameter estimation. Estimates for ϕ and σ^2 can be obtained by plugging in the sample autocovariances $\hat{\gamma}_0, \hat{\gamma}_1, \ldots \hat{\gamma}_p$, and solving the Y-W equations.

Parameter estimation for ARMA processes

- Estimation using Yule-Walker equations is a **method of moments (MoM)** approach to parameter estimation. Note that we do not use the MoM approach to estimate MA processes, or the MA component of ARMA processes, because of instability of the resulting estimates (parameters are highly nonlinear functions of the data).
- **Least squares** (conditional and unconditional) estimation;
- **Maximum likelihood** estimation.

Specification: order determination

- ACF of MA processes;
- Partial autocorrelation function (PACF) of AR processes;
- Extended partial autocorrelation function (EACF) of ARMA processes;
- Unit-root nonstationarity;
- Selection rules based on AIC and BIC.

Model specification is always tentative and subject to further examination, diagnostic checking, and modification if necessary.

Sample ACF properties

Sample autocorrelation function (sample ACF):

$$\hat{\rho}_k = r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}.$$

Theorem (Bartlett): sampling distribution of $(r_1, \ldots, r_k)$

Let $Y_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j}$ be a stationary general linear process.
Then for fixed m and as $n \to \infty$,

$$\sqrt{n}\left((r_1 - \rho_1), \ldots, (r_m - \rho_m)\right) \overset{approx}{\sim} N\left(\mathbf{0}, \mathbf{C} = \{c_{ij}\}\right)$$

where $c_{ij} = \sum_{k=-\infty}^{\infty} \left(\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k^2\right).$

For n large,

- $r_k \overset{approx}{\sim} N(\rho_k, c_{kk}/n)$ $\quad \to Var(r_k) = O(1/n)$;
- $Cor(r_k, r_j) \approx c_{kj}/\sqrt{c_{jj}c_{kk}}$ $\quad \to Cor(r_k, r_j) = O(1)$!

Sample ACF properties: special cases

$\{Y_t\}$ is white noise:

- $r_k \overset{approx}{\sim} N(0, 1/n)$ for large n, and
- $r_k \perp r_j$ for all $k \neq j$.

$\{Y_t\}$ is AR(1):

- $\rho_k = \phi^k$ for all k;
- $Var(r_1) \approx (1-\phi^2)/n$: $Var(r_1) \to 0$ as $|\phi| \to 1$;
- for k large, $Var(r_k) \approx [(1+\phi^2)/(1-\phi^2)]/n$: $Var(r_k) \to \infty$ as $|\phi| \to 1$!
- $Cor(r_1, r_2) \approx 2\phi\sqrt{(1-\phi^2)/(1+2\phi^2-3\phi^4)}$.

$\{Y_t\}$ is MA(1):

- $\rho_1 = -\frac{\theta}{1+\theta^2}$;
- $Var(r_1) \approx (1-3\rho_1^2+4\rho_1^4)/n$;
- for $k > 1$, $Var(r_k) \approx (1+2\rho_1^2)/n$: $Var(r_k) > Var(r_1)$
- $Cor(r_1, r_2) \approx 2\rho_1(1-\rho_1^2)$.

$\{Y_t\}$ is MA(q):

- for $k > q$, $Var(r_k) \approx (1+2\sum_{j=1}^{q}\rho_j^2)/n$.

Estimated variance of r_k

Suppose the null hypothesis is that $\{Y_t\}$ follows MA(1) process with $\theta = -0.7$. Under the null, with $n = 100$:

- $\rho_1 = -\frac{\theta}{1+\theta^2} \approx 0.47$;
- $Var(r_1) \approx (1 - 3\rho_1^2 + 4\rho_1^4)/n = 0.00532$;
- $Var(r_2) \approx (1 + 2\rho_1^2)/n = 0.0144$;
- $Cor(r_1, r_2) \approx 0.83$.

The null model would be rejected at $\alpha = 5\%$ if for $k \geq 2$,

$$\frac{r_k}{\sqrt{Var(r_k)}} > 1.96,$$

or equivalently, $|r_k| > 0.24$.
The problem is, we do not know the true process hence the true ρ_1, and need to use an estimated for ρ_1 (that is, r_1) to estimate the variance of r_k!

Estimated variance of r_k

Under the null model (MA(1) with $\theta = -0.7$), with $n = 100$:

- $\rho_1 = -\frac{\theta}{1+\theta^2} \approx 0.47$;
- $Var(r_1) \approx (1 - 3\rho_1^2 + 4\rho_1^4)/n = 0.0532$;
- $Var(r_2) \approx (1 + 2\rho_1^2)/n = 0.0144$;
- $Cor(r_1, r_2) \approx 0.83$.

95% CI ρ_k:

$(r_k \pm 1.96 \cdot \sqrt{Var(r_k)})$

$k=1$ estimates

With observed $r_1 = 0.32$:

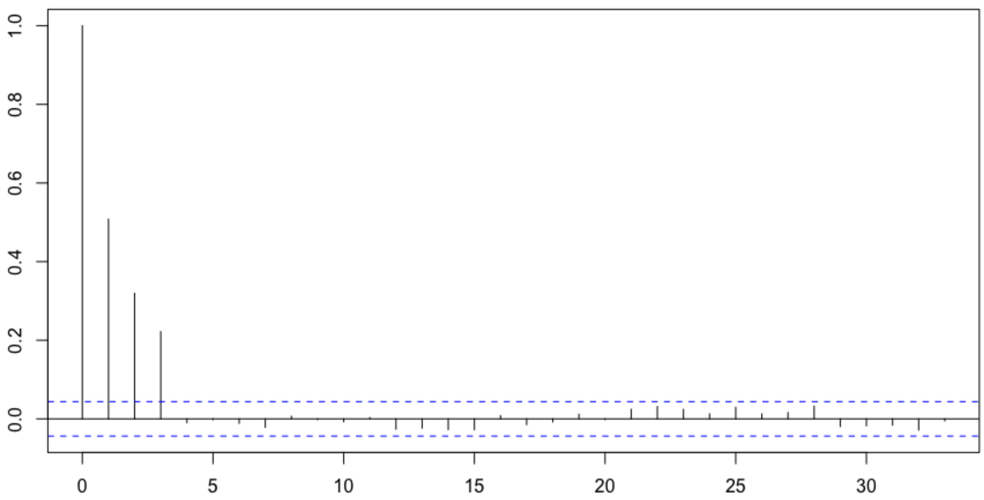
- $\hat{Var}(r_1) \approx (1 - 3r_1^2 + 4r_1^4)/n = 0.00735$;
- $\hat{Var}(r_k) \approx (1 + 2r_1^2)/n = 0.012$ for $k \geq 2$.

$(0.32 \pm 1.96\sqrt{0.00735})$

Question: how to construct 95% confidence intervals for ρ_1 and ρ_k ($k \geq 2$)? If the null model were true, what values of ρ_1 and ρ_k should these intervals contain 95% of the time?

ACF for MA order determination

- Assume $\{Y_t\}$ is a MA(q) process.
- Then, $\rho_k = 0$ for $k > q$, that is the theoretical ACF cuts off after q. We explore the sample ACF to determine q. There are two cases:
  - $r_q \xrightarrow{p} \rho_q \neq 0$.
  - When $k > q$, $r_k \xrightarrow{p} 0$, and $\sqrt{n} r_k \Rightarrow N(0, 1 + 2(\rho_1^2 + \cdots + \rho_q^2))$.
- The two blue lines are at $\pm z_{0.975}/\sqrt{n}$ respectively.



PACF of stationary time series

Suppose we want to predict Y_t and Y_{t-k} using $\{Y_{t-1}, \ldots, Y_{t-k+1}\}$ via OLS.

$$
\begin{aligned}
\hat{Y}_t &= \beta_1 Y_{t-1} + \cdots + \beta_{k-1} Y_{t-k+1} \\
\hat{Y}_{t-k} &= \beta_1 Y_{t-k+1} + \cdots + \beta_{k-1} Y_{t-1}
\end{aligned}
$$

The lag-k partial autocorrelation function (PACF):

$$\phi_{kk} = Cor\left(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}\right).$$

That is, ϕ_{kk} is the correlation between Y_t and Y_{t-k} after removing the effect of intervening variables $\{Y_{t-1}, \ldots, Y_{t-k+1}\}$.

- If $\{Y_t\}$ are jointly normal, then $\phi_{kk} = Cor\left(Y_t, Y_{t-k} \mid Y_{t-1}, \ldots, Y_{t-k+1}\right)$;
- $\phi_{11} = \rho_1$; (why?)
- $\phi_{22} = \left(\rho_2 - \rho_1^2\right) / \left(1 - \rho_1^2\right)$, using the fact that $\hat{Y}_t = \rho_1 Y_{t-1}$;

PACF of stationary time series

$$\phi_{11} = \rho_1, \qquad \phi_{22} = \left(\rho_2 - \rho_1^2\right) / \left(1 - \rho_1^2\right)$$

- AR(1): $Y_t = \phi Y_{t-1} + e_t$

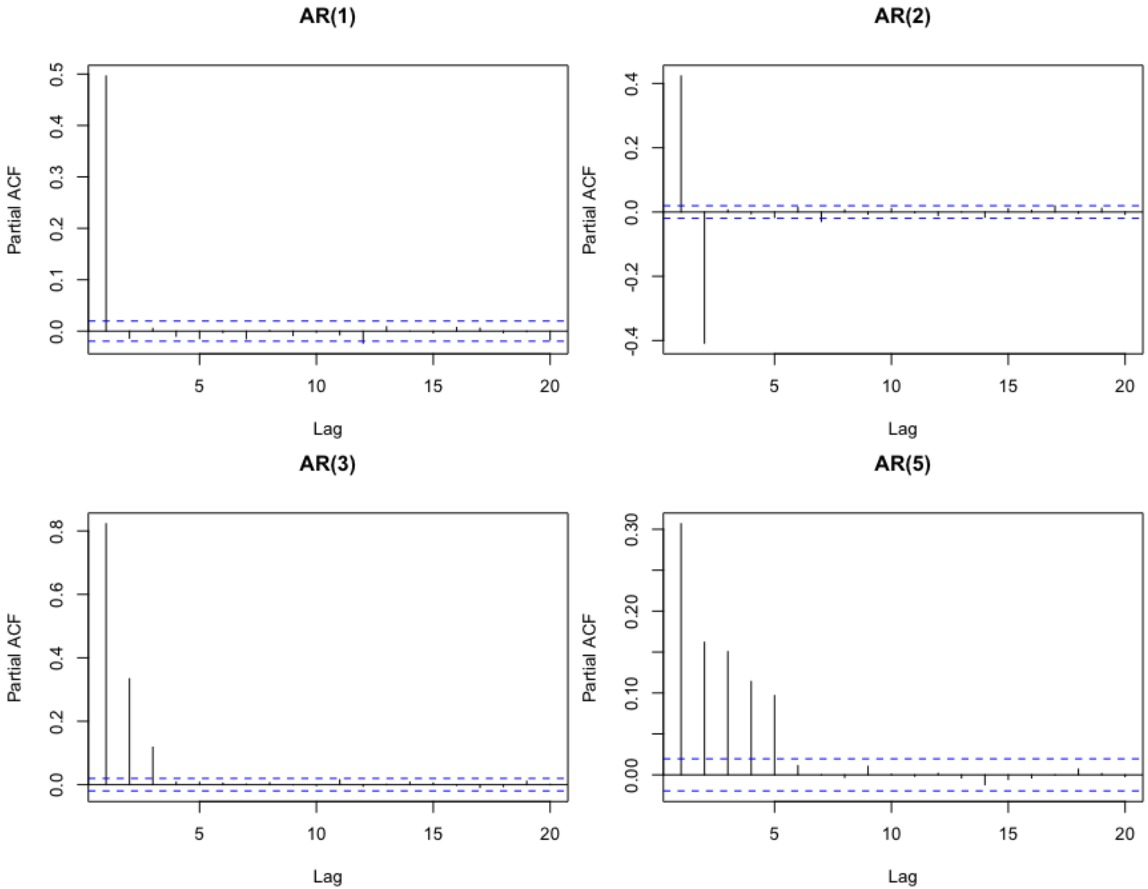
$$\phi_{22} = (\phi^2 - \phi^2)/(\phi^2 - \phi^2) = 0.$$

  In fact, $\phi_{kk} = 0$ for all $k > 1$.
- AR(p): $\phi_{kk} = 0$ for all $k > p$. That is, the PACF of an AR process cuts off after its order.
- MA(1): $Y_t = e_t - \theta e_{t-1}$. For $k \geq 1$,

$$\phi_{kk} = -\frac{\theta^k \left(1 - \theta^2\right)}{1 - \theta^{2(k+1)}},$$

  which decays (nearly) exponentially, like the ACF of AR(1).
- MA(q): PACF behaves similarly to the ACF of AR(q).

Examples of PACF of AR processes



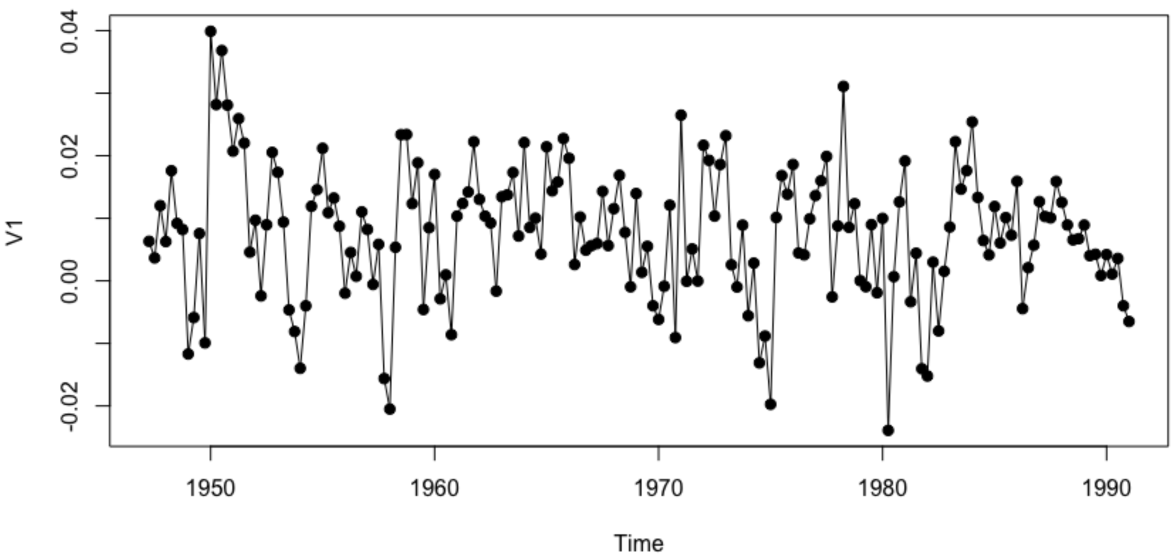
PACF for AR order determination

- The PACF ϕ_{kk} can be estimated by solving the k-th order Yule-Walker equations. The estimate $\hat{\phi}_{kk}$ is called sample PACF of lag-k.
- Assume the process is AR(p). Depending on $k = p$ or $k > p$, there are two cases:
  (i) $\hat{\phi}_{pp} \to \phi_p$ in probability as $n \to \infty$.
  (ii) When $k > p$, $\hat{\phi}_{kk} \xrightarrow{p} 0$. Furthermore, $\sqrt{n}\hat{\phi}_{kk} \Rightarrow N(0, 1)$.
- Since the theoretical PACF cuts off after p, we explore the sample PACF to determine p.
- Suppose we wish to test $H_0 : \ \phi_{kk} = 0$ vs $H_0 : \ \phi_{kk} \neq 0$. A level α test should reject H_0 if $|\hat{\phi}_{kk}| > z_{1-\alpha/2}/\sqrt{n}$.
- Choose a level α, pick the AR order as the smallest integer p^* such that $|\hat{\phi}_{kk}| \leq z_{1-\alpha/2}/\sqrt{n}$ for all $k > p^*$.

Example: U.S. GNP

Quarterly growth rate of U.S. real gross national product (GNP), from the second quarter of 1947 to the first quarter of 1991.

```
gnp=fread('http://faculty.chicagobooth.edu/ruey.tsay/teaching/
           fts3/dgnp82.txt')
gnp1=ts(gnp,frequency=4,start=c(1947,2))
plot(gnp1); points(gnp1,pch=19)
```



Example: U.S. GNP

Fitting an AR(3) model (order chosen by AIC):

```
gnp.ar3=arima(gnp,order=c(3,0,0)); gnp.ar3
## Coefficients:
##          ar1     ar2      ar3  intercept
##       0.3480  0.1793  -0.1423     0.0077
## s.e.  0.0745  0.0778   0.0745     0.0012
##
## sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,
   aic = -1121.68
```

σ μ

The constant term is obtained as $(1 - .348 - .1793 + .1423) * 0.0077$.
The residual standard error is

```
sqrt(gnp.ar3$sigma2)
## [1] 0.009709322
```

Testing $\phi_3 = 0$: reject null at 5% level if absolute value > 1.96:

```
-0.1423/0.0745
## [1] -1.910067
```

EACF of ARMA processes

- For ARMA processes, neither the ACF or PACF cuts off after the true order, making identification harder.
- The extended autocorrelation function (EACF) aims to help with identifying ARMA orders.
- The idea is, if the AR process of an ARMA model is known, then after "filtering" it out from the observed series (these coefficients can be estimated with a finite sequence of regressions), the remainder part is purely MA.

EACF of ARMA processes

When reading the EACF table, look for a triangular pattern of O's. Its upper left vertex will identify the order of (p, q).

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

Keep in mind that the sample EACF will never be this clean-cut.

Overdifferencing

Example. Let $\{Y_t\}$ be the random walk.

- First difference $\nabla Y_t = e_t$ is white noise: stationarity is achieved.
- Second difference $\nabla^2 Y_t = e_t - e_{t-1}$ is still stationary, however, it is a non-invertible MA process.

Keep in mind the principle of parsimony: preference towards the least complex explanation for the observation.

Unit root nonstationarity

ARIMA(p,1,q): $W_t = \nabla Y_t$, where

$$\begin{aligned} W_t &= \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}. \end{aligned}$$

Solving for Y_t, we have

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

Why isn't this an ARMA(p+1, q) model?

Unit root nonstationarity

ARIMA(p,1,q): $W_t = \nabla Y_t$, where

$$\begin{aligned} W_t &= \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t \\ &\quad -\theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}. \end{aligned}$$

Solving for Y_t, we have

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

Why isn't this an ARMA(p+1, q) model? CP of the AR component:

$$\begin{aligned} \phi(z) &= 1 - (1+\phi_1) z - (\phi_2 - \phi_1) z^2 - \cdots - \left(\phi_p - \phi_{p-1}\right) z^p + \phi_p z^{p+1} \\ &= \left(1 - \phi_1 z^2 - \cdots - \phi_p z^p\right)(1 - z) \end{aligned}$$

for which $z = 1$ is a root: nonstationarity!

Dickey-Fuller test of unit root

- If the sample ACF decays linearly, it is a sign of unit root (nonstationarity).
- Idea: suppose

$$
\begin{aligned}
Y_t &= \alpha Y_{t-1} + X_t, \\
X_t &= \phi_1 X_1 + \cdots + \phi_k X_k + e_t \ \text{ is AR(k).}
\end{aligned}
$$

- Then, we have

$$\nabla Y_t = \underbrace{(\alpha - 1)}_{\mathbf{a}} Y_{t-1} + \phi_1 \nabla Y_{t-1} + \cdots + \phi_k \nabla Y_{t-k} + e_t$$

- Under $H_0 : \mathbf{a} = (\alpha - 1) = 0$, $X_t = \nabla Y_t$, and Y_t is nonstationary.
- Equivalently, $\{Y_t\}$ is an AR(p+1) process with characteristic equation

$$\phi(z)(1 - \alpha z) = 0,$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_k z^k$, and H_0 amounts to testing whether this AR(p+1) process has a unit root.

Dickey-Fuller test of unit root

$$\nabla Y_t = \underbrace{(\alpha - 1)}_{\mathbf{a}} Y_{t-1} + \phi_1 \nabla Y_{t-1} + \cdots + \phi_k \nabla Y_{t-k} + e_t$$

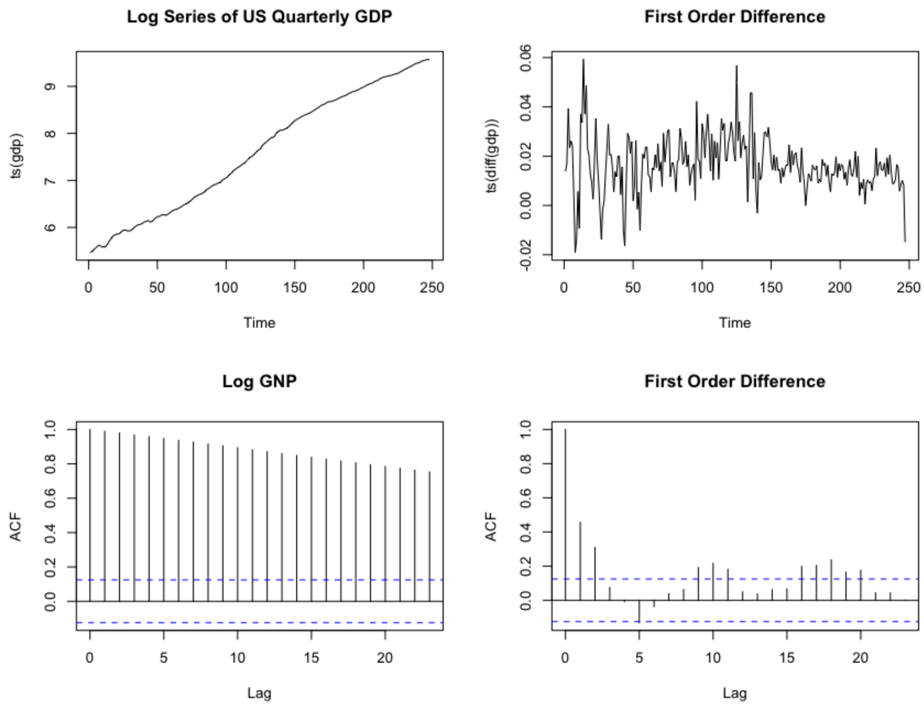
- Let $\hat{\mathbf{a}}$ be the least squares regression estimate of $\mathbf{a}$. To test

$$H_0 : \ \mathbf{a} = 0 \ \text{ vs } \ H_1 : \ \mathbf{a} < 0,$$

  the augmented Dickey-Fuller (ADF) test uses the t-ratio $\hat{\mathbf{a}}/\text{sd}(\hat{\mathbf{a}})$ as the test statistic.
- The reference distribution of the ADF test is not approximately t-distributed under the null. However, we can get numerical values from R.

Example: US Quarterly GDP from 1947 to 2008

```
library(fUnitRoots)
da=fread('http://faculty.chicagobooth.edu/ruey.tsay/teaching/
          fts3/q-gdp4708.txt')
gdp=log(da[,4])
plot(ts(gdp), main="Log Series of US Quarterly GDP")
plot(ts(diff(gdp)), main="First Order Difference")
acf(gdp,main="Log GNP")
acf(diff(gdp),main="First Order Difference")
```



Example: US Quarterly GDP from 1947 to 2008

```
gdp.d.ar=ar(diff(gdp), method="mle")
gdp.d.ar$order
## [1] 10

adfTest(gdp, lag=10, type=c("c"))

## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 10
##   STATISTIC:
##     Dickey-Fuller: -1.6109
##   P VALUE:
##     0.4569
```

*Note: `type` describes the type of the unit root regression: "nc" for a regression with no intercept (constant) nor time trend; "c" for a regression with an intercept (constant) but no time trend; "ct" for a regression with an intercept (constant) and a time trend. The default is "c".

Order specification using AIC and BIC

- Akaike information criterion:

$$\text{AIC} = -2\log(\text{maximum likelihood}) + 2k$$

  where $k = p + q + 1$ if the model contains a constant term, and $k = p + q$ otherwise.
- Bayesian information criterion:

$$\text{BIC} = -2\log(\text{maximum likelihood}) + k\log n$$

- Select ARMA orders based on minimizing AIC and BIC.
  - If the true process is ARMA(p, q), order specification by minimizing BIC is consistent as $n \to \infty$;
  - If the true process is not ARMA(p, q), order specification by minimizing AIC leads to a "closest" (in terms of KL divergence) model to the true process among a growing pool of models.
- BIC tends to select a smaller order, due to heavier weights on the number of parameters.

Example: U.S. GNP

AR order determination by AIC for the U.S. GNP data:

```
gnp.pacf=acf(gnp,type="partial")
gnp.ord=ar(gnp,method='yw',ord.max=30); gnp.ord$order
## [1] 3
gnp.ord$aic ## has been adjusted so that the minimum AIC is zero
##          0          1          2          3          4          5
## 27.5691310  2.6081086  1.5895550  0.0000000  0.2734771  2.2034466
##          6          7          8          9         10         11
##  4.0171066  5.9916210  5.8264833  7.5230025  7.8223499  9.5813222
```