Chapter 2: Linear Time Series (TS) Models

Financial TS: collection of a financial measurement over time
Example: log return r_t

Data: $\{r_1, r_2, \cdots, r_T\}$ (T data points)
Purpose: What is the information contained in $\{r_t\}$?

Basic concepts

- Stationarity:
  - Strict: distributions are time-invariant
  - Weak: first 2 moments are time-invariant

  What does weak stationarity mean in practice?

  Past: time plot of $\{r_t\}$ varies around a fixed level within a finite range!

  Future: the first 2 moments of future r_t are the same as those of the data so that meaningful inferences can be made.

- Mean (or expectation) of returns:

$$\mu = E(r_t)$$

- Variance (variability) of returns:

$$\mathrm{Var}(r_t) = E[(r_t - \mu)^2]$$

- Sample mean and sample variance are used to estimate the mean and variance of returns.

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t \quad \& \quad \mathrm{Var}(r_t) = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})^2$$

- Test $H_o : \mu = 0$ vs $H_a : \mu \neq 0$. Compute
$$t = \frac{\bar{r}}{\text{std}(\bar{r})} = \frac{\bar{r}}{\sqrt{\text{Var}(r_t)/T}}$$
Compare t ratio with $N(0,1)$ dist.
Decision rule: Reject H_o of zero mean if $|t| > Z_{\alpha/2}$ or p-value is less than α.

- Lag-k autocovariance:
$$\gamma_k = \text{Cov}(r_t, r_{t-k}) = E[(r_t - \mu)(r_{t-k} - \mu)].$$

- Serial (or auto-) correlations:
$$\rho_\ell = \frac{\text{cov}(r_t, r_{t-\ell})}{\text{var}(r_t)}$$
Note: $\rho_0 = 1$ and $\rho_k = \rho_{-k}$ for $k \neq 0$. Why?
Existence of serial correlations implies that the return is predictable, indicating market inefficiency.

- Sample autocorrelation function (ACF)
$$\hat{\rho}_\ell = \frac{\Sigma_{t=1}^{T-\ell}(r_t - \bar{r})(r_{t+\ell} - \bar{r})}{\Sigma_{t=1}^{T}(r_t - \bar{r})^2},$$
where $\bar{r}$ is the sample mean & T is the sample size.

- Test zero serial correlations (market efficiency)
  - Individual test: for example,
    $H_o : \rho_1 = 0$ vs $H_a : \rho_1 \neq 0$
$$t = \frac{\hat{\rho}_1}{\sqrt{1/T}} = \sqrt{T}\hat{\rho}_1$$

Asym. $N(0,1)$.
Decision rule: Reject H_o if $|t| > Z_{\alpha/2}$ or p-value less than α.

- Joint test (Ljung-Box statistics):
  $H_o : \rho_1 = \cdots = \rho_m = 0$ vs $H_a : \rho_i \neq 0$

$$Q(m) = T(T+2) \sum_{\ell=1}^{m} \frac{\hat{\rho}_\ell^2}{T-\ell}$$

  Asym. chi-squared dist with m degrees of freedom.
  Decision rule: Reject H_o if $Q(m) > \chi_m^2(\alpha)$ or p-value is less than α.

- Sources of serial correlations in financial TS
  - Nonsynchronous trading
  - Bid-ask bounce
  - Risk premium, etc.

  Thus, significant sample ACF does not necessarily imply market inefficiency.

Example: Monthly returns of IBM stock from 1926 to 1997.

- R_t: $Q(5) = 5.4(0.37)$ and $Q(10) = 14.1(0.17)$
- r_t: $Q(5) = 5.8(0.33)$ and $Q(10) = 13.7(0.19)$

Remark: What is p-value? How to use it?
Implication: Monthly IBM stock returns do not have significant serial correlations.
Example: Monthly returns of CRSP value-weighted index from 1926 to 1997.

- R_t: $Q(5) = 27.8$ and $Q(10) = 36.0$
- r_t: $Q(5) = 26.9$ and $Q(10) = 32.7$

All highly significant. Implication: there exist significant serial correlations in the value-weighted index returns. (Nonsynchronous trading might explain the existence of the serial correlations, among other reasons.) Similar result is also found in equal-weighted index returns.

R demonstration: IBM monthly simple returns from 1968 to 2015

```
> da=read.table("m-ibm-6815.txt",header=T)
> ibm=da$RET
> acf(ibm) %% Plot not shown
> m1 <- acf(ibm)
> names(m1)
[1] "acf"    "type"   "n.used" "lag"    "series" "snames"
> m1$acf
               [,1]
 [1,]  1.0000000000  % lag 0
 [2,] -0.0068713539  % lag 1
 [3,] -0.0002212888
 ....
[28,]  0.0159729906
```

The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.

```
> m2 <- pacf(ibm)  % Partial ACF
> names(m2)
[1] "acf"    "type"   "n.used" "lag"    "series" "snames"
> m2$acf
               [,1]
 [1,]  -0.0068713539
 [2,]  -0.0002685169
 [3,]   0.0310623477
 ....
[27,]   0.0127614307

> Box.test(ibm,lag=10)  % Box-Pierce Q(m) test
        Box-Pierce test
data:  ibm
X-squared = 7.1714, df = 10, p-value = 0.7092

> Box.test(ibm,lag=10,type='Ljung') % Ljung-Box Q(m) test
        Box-Ljung test
data:  ibm
```

```
X-squared = 7.2759, df = 10, p-value = 0.6992
```

Back-shift (lag) operator
A useful notation in TS analysis.

- Definition: $Br_t = r_{t-1}$ or $Lr_t = r_{t-1}$
- $B^2 r_t = B(Br_t) = Br_{t-1} = r_{t-2}$.

B (or L) means time shift! Br_t is the value of the series at time $t-1$.
Suppose that the daily log returns are

Day	1	2	3	4
r_t	0.017	-0.005	-0.014	0.021

Answer the following questions:

- $r_2 =$
- $Br_3 =$
- $B^2 r_5 =$

Question: What is $B2$?

What are the important statistics in practice?
Conditional quantities, not unconditional

A proper perspective: at a time point t

- Available data: $\{r_1, r_2, \cdots, r_{t-1}\} \equiv F_{t-1}$
- The return is decomposed into two parts as

$$\begin{aligned} r_t &= \text{predictable part} + \text{not predictable part} \\ &= \text{function of elements of } F_{t-1} + a_t \end{aligned}$$

In other words, given information F_{t-1}

$$\begin{aligned} r_t &= \mu_t + a_t \\ &= E(r_t|F_{t-1}) + \sigma_t \epsilon_t \end{aligned}$$

- μ_t: conditional mean of r_t
- a_t: shock or innovation at time t
- ϵ_t: an iid sequence with mean zero and variance 1
- σ_t: conditional standard deviation (commonly called volatility in finance)

Traditional TS modeling is concerned with μ_t:
Model for μ_t: **mean equation**
Volatility modeling concerns σ_t.
Model for σ_t^2: **volatility equation**

Univariate TS analysis serves two purposes

- a model for μ_t
- understanding models for σ_t^2: properties, forecasting, etc.

Linear time series: r_t is linear if

- the predictable part is a linear function of F_{t-1}
- $\{a_t\}$ are independent and have the same dist. (iid)

Mathematically, it means r_t can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i},$$

where μ is a constant, $\psi_0 = 1$ and $\{a_t\}$ is an iid sequence with mean zero and well-defined distribution.

In the economic literature, a_t is the *shock* (or *innovation*) at time t and $\{\psi_i\}$ are the *impulse* responses of r_t.

White noise: iid sequence (with finite variance), which is the building block of linear TS models.
White noise is not predictable, but has zero mean and finite variance.

Univariate linear time series models

1. autoregressive (AR) models
2. moving-average (MA) models
3. mixed ARMA models

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

An AR(3) model for the data is

$$r_t = 0.005 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.01,$$

where $\{a_t\}$ denotes a white noise with variance σ_a^2. Given r_n, r_{n-1} & r_{n-2}, we can predict r_{n+1} as

$$\hat{r}_{n+1} = 0.005 + 0.35r_n + 0.18r_{n-1} - 0.14r_{n-2}.$$

Other implications of the model?

In this course, we use *statistical methods* to find models that fit the data well for making inference, e.g. prediction. On the other hand, there exists economic theory that leads to time-series models for economic variables. For instance, consider the *real business-cycle theory* in macroeconomics. Under some simplifying assumptions, one can show that $\ln(Y_t)$, where Y_t is the output (GDP), follows an AR(2) model. See *Advanced Macroeconomics* by David Romer (2006, 3rd, pp. 190).

Example: Monthly simple return of Center for Research in Security Prices (CRSP) equal-weighted index

$$R_t = 0.013 + a_t + 0.178a_{t-1} - 0.13a_{t-3} + 0.135a_{t-9}, \quad \hat{\sigma}_a = 0.073$$

Checking: $Q(10) = 11.4(0.122)$ for the residual series a_t.
Implications of the model?
Statistical significance vs economic significance.

In this course, we shall discuss some reasons for the observed serial dependence in index returns. See, for example, Chapter 5 on nonsynchronous trading.

Important properties of a model

- Stationarity condition
- Basic properties: mean, variance, serial dependence
- Empirical model building: specification, estimation, & checking
- Forecasting

Simple AR models: (Regression with lagged variables.)

Motivating example: The growth rate of U.S. quarterly real GNP from 1947 to 1991. Recall that the model discussed before is

$$r_t = 0.005 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \hat{\sigma}_a = 0.01.$$

This is called an AR(3) model because the growth rate r_t depends on the growth rates of the past **three** quarters. How do we specify this model from the data? Is it adequate for the data? What are the implications of the model? These are the questions we shall address in this lecture.

Another example: U.S. monthly unemployment rate.

<u>AR(1) model</u>:

1. Form: $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, where ϕ_0 and ϕ_1 are real numbers, which are referred to as "parameters" (to be estimated from the data in an application). For example,

$$r_t = 0.05 + 0.4r_{t-1} + a_t$$

2. Stationarity: necessary and sufficient condition $|\phi_1| < 1$. Why?

3. Mean: $E(r_t) = \frac{\phi_0}{1-\phi_1}$

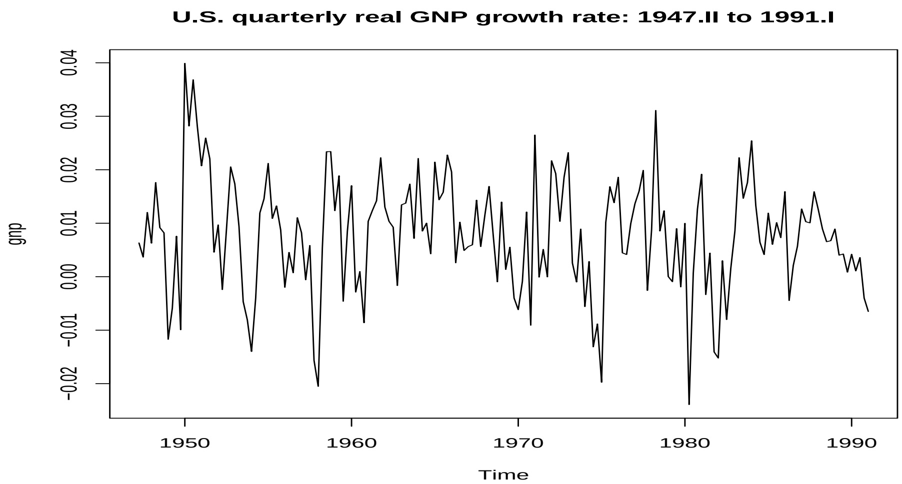


Figure 1: U.S. quarterly growth rate of real GNP: 1947-1991

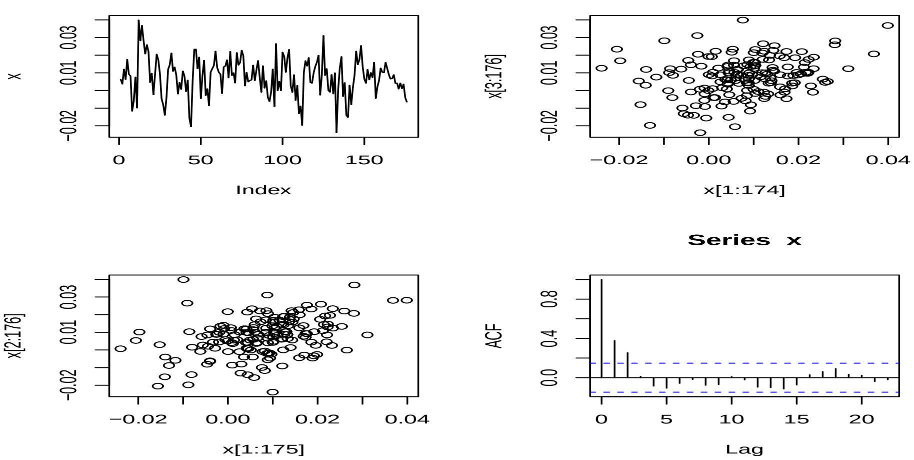


Figure 2: Various plots of U.S. quarterly growth rate of real GNP: 1947-1991

Figure 3: U.S. monthly unemployment rate (total civilian, 16 and older) from January 1948 to February, 2017.

4. Alternative representation: Let $E(r_t) = \mu$ be the mean of r_t so that $\mu = \phi_0/(1 - \phi_1)$. Equivalently, $\phi_0 = \mu(1 - \phi_1)$. Plugging in the model, we have

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + a_t. \tag{1}$$

   This model also has two parameters (μ and ϕ_1). It explicitly uses the mean of the series. It is less commonly used in the literature, but is the model representation used in R.

5. Variance: $\mathrm{Var}(r_t) = \frac{\sigma_a^2}{1-\phi_1^2}$.

6. Autocorrelations: $\rho_1 = \phi_1, \rho_2 = \phi_1^2$, etc. In general, $\rho_k = \phi_1^k$ and ACF ρ_k decays exponentially as k increases,

7. Forecast (minimum squared error): Suppose the forecast origin is n. For simplicity, we shall use the model representation in (1)

and write $x_t = r_t - \mu$. The model then becomes $x_t = \phi_1 x_{t-1} + a_t$. Note that forecast of r_t is simply the forecast of x_t plus μ.

(a) 1-step ahead forecast at time n:

$$\hat{x}_n(1) = \phi_1 x_n$$

(b) 1-step ahead forecast error:

$$e_n(1) = x_{n+1} - \hat{x}_n(1) = a_{n+1}$$

Thus, a_{n+1} is the *un-predictable* part of x_{n+1}. It is the shock at time $n+1$!

(c) Variance of 1-step ahead forecast error:

$$\mathrm{Var}[e_n(1)] = \mathrm{Var}(a_{n+1}) = \sigma_a^2.$$

(d) 2-step ahead forecast:

$$\hat{x}_n(2) = \phi_1 \hat{x}_n(1) = \phi_1^2 x_n.$$

(e) 2-step ahead forecast error:

$$e_n(2) = x_{n+2} - \hat{x}_n(2) = a_{n+2} + \phi_1 a_{n+1}$$

(f) Variance of 2-step ahead forecast error:

$$\mathrm{Var}[e_n(2)] = (1 + \phi_1^2)\sigma_a^2$$

which is greater than or equal to $\mathrm{Var}[e_n(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.

(g) Behavior of multi-step ahead forecasts. In general, for the ℓ-step ahead forecast at n, we have

$$\hat{x}_n(\ell) = \phi_1^\ell x_n,$$

the forecast error

$$e_n(\ell) = a_{n+\ell} + \phi_1 a_{n+\ell-1} + \cdots + \phi_1^{\ell-1} a_{n+1},$$

and the variance of forecast error

$$\mathrm{Var}[e_n(\ell)] = (1 + \phi_1^2 + \cdots + \phi_1^{2(\ell-1)})\sigma_a^2.$$

In particular, as $\ell \to \infty$,

$$\hat{x}_n(\ell) \to 0, \quad i.e., \quad \hat{r}_n(\ell) \to \mu.$$

This is called the *mean-reversion* of the AR(1) process. The variance of forecast error approaches

$$\mathrm{Var}[e_n(\ell)] = \frac{1}{1-\phi_1^2}\sigma_a^2 = \mathrm{Var}(r_t).$$

In practice, it means that for the long-term forecasts serial dependence is not important. The forecast is just the sample mean and the uncertainty is simply the uncertainty about the series.

8. A compact form: $(1 - \phi_1 B) r_t = \phi_0 + a_t$.

Half-life: A common way to quantify the *speed* of mean reversion is the half-life, which is defined as the number of periods needed so that the magnitude of the forecast becomes half of that of the forecast origin. For an AR(1) model, this mean

$$x_n(k) = \frac{1}{2}x_n.$$

Thus, $\phi_1^k x_n = \frac{1}{2}x_n$. Consequently, the half-life of the AR(1) model is $k = \frac{\ln(0.5)}{\ln(|\phi_1|)}$. For example, if $\phi_1 = 0.5$, the $k = 1$. If $\phi_1 = 0.9$, then $k \approx 6.58$.

<u>AR(2) model</u>:

1. Form: $r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t$, or

$$(1 - \phi_1 B - \phi_2 B^2) r_t = \phi_0 + a_t.$$

2. Stationarity condition: (factor of polynomial)

3. Characteristic equation: $(1 - \phi_1 x - \phi_2 x^2) = 0$

4. Mean: $E(r_t) = \frac{\phi_0}{1-\phi_1-\phi_2}$

5. Mean-adjusted format: Using $\phi_0 = \mu - \phi_1\mu - \phi_2\mu$, we can write the AR(2) model as

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t.$$

   This form is often used in the finance literature to highlight the mean-reverting property of a stationary AR(2) model.

6. ACF: $\rho_0 = 1$, $\rho_1 = \frac{\phi_1}{1-\phi_2}$,

$$\rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-1}, \quad \ell \geq 2.$$

7. Stochastic business cycle: if $\phi_1^2 + 4\phi_2 < 0$, then r_t shows characteristics of business cycles with average length
$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]},$$
where the cosine inverse is stated in radian. If we denote the solutions of the polynomial as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a$ and $\phi_2 = -(a^2 + b^2)$ so that
$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2 + b^2})}.$$
In R or S-Plus, one can obtain $\sqrt{a^2 + b^2}$ using the command **Mod**.

8. Forecasts: Similar to AR(1) models

Simulation in R: Use the command `arima.sim`

1. `y1=arima.sim(model=list(ar=c(1.3,-.4)),1000)`

2. `y2=arima.sim(model=list(ar=c(.8,-.7)),1000)`

Check the ACF and PACF of the above two simulated series.

Discussion: (Reference only)
An AR(2) model can be written as an AR(1) model if one expands the dimension. Specifically, we have

$$\begin{aligned} r_t - \mu &= \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t \\ r_{t-1} - \mu &= r_{t-1} - \mu, \quad \text{(an identity.)} \end{aligned}$$

Now, putting the two equations together, we have

$$\begin{bmatrix} r_t - \mu \\ r_{t-1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} - \mu \\ r_{t-2} - \mu \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}.$$

This is a 2-dimensional AR(1) model. Several properties of the AR(2) model can be obtained from the expanded AR(1) model.

Building an AR model

- Order specification

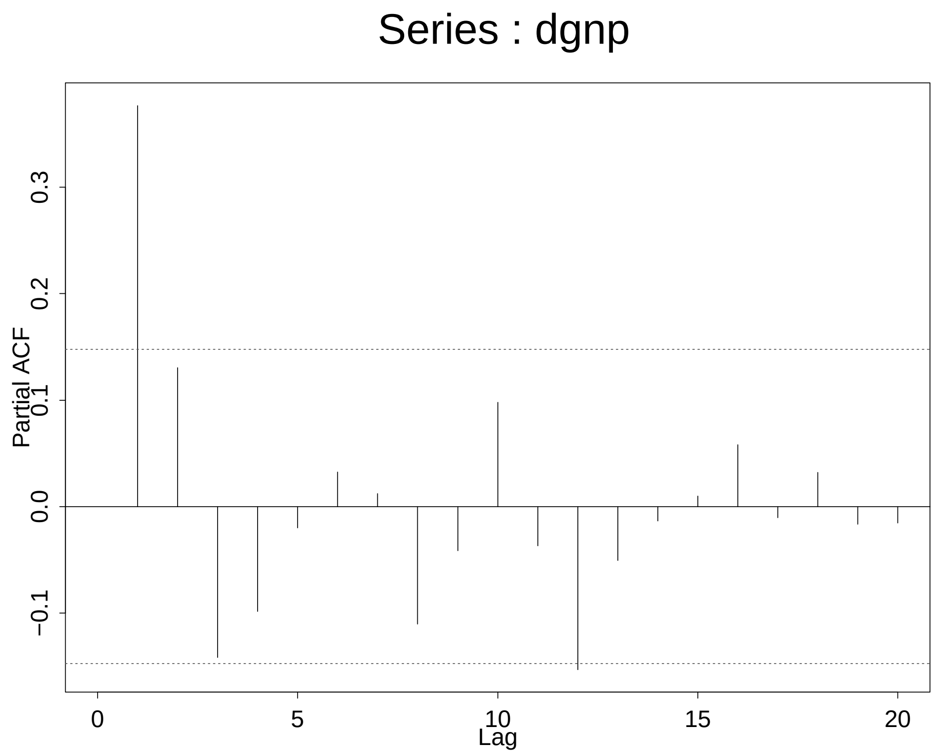
  1. Partial ACF: (naive, but effective)
     - Use consecutive fittings
     - See Text (p. 40) for details
     - **Key feature**: PACF cuts off at lag p for an AR(p) model.
     - Illustration: See the PACF of the U.S. quarterly growth rate of GNP.
  2. Akaike information criterion

     $$AIC(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{2\ell}{T},$$

     for an AR(ℓ) model, where $\tilde{\sigma}_\ell^2$ is the MLE of residual variance.

     Find the AR order with *minimum* AIC for $\ell \in [0, \cdots, P]$.
  3. BIC criterion:

     $$BIC(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{\ell \ln(T)}{T}.$$



R command: `ar(rt, method=''mle'',order.max=12)`

- Needs a constant term? Check the sample mean.
- Estimation: least squares method or maximum likelihood method
- Model checking:

1. Residual: obs minus the fit, i.e. 1-step ahead forecast errors at each time point.
2. Residual should be close to white noise if the model is adequate. Use Ljung-Box statistics of residuals, but degrees of freedom is $m - g$, where g is the number of AR coefficients used in the model.

Example: Analysis of U.S. GNP growth rate series.
R demonstration:

```
> setwd("your working directory")
> library(fBasics)
> da=read.table("dgnp82.dat")
> x=da[,1]
> par(mfcol=c(2,2))  % put 4 plots on a page ###
See Figure 2 of the lecture note 2.
> plot(x,type='l')  % first plot
> plot(x[1:175],x[2:176]) % 2nd plot
> plot(x[1:174],x[3:176]) % 3rd plot
> acf(x,lag=12)  % 4th plot

> pacf(x,lag.max=12) % Compute PACF (not shown in this handout)
> Box.test(x,lag=10,type='Ljung') % Compute Q(10) statistics
        Box-Ljung test
data:  x
X-squared = 43.2345, df = 10, p-value = 4.515e-06

> m1=ar(x,method='mle') % Automatic AR fitting using AIC criterion.
> m1
Call: ar(x = x, method = "mle")
Coefficients:
      1        2        3      % An AR(3) is specified.
 0.3480   0.1793  -0.1423

Order selected 3  sigma^2 estimated as  9.427e-05

> names(m1)
 [1] "order"        "ar"           "var.pred"     "x.mean"       "aic"
 [6] "n.used"       "order.max"    "partialacf"   "resid"        "method"
[11] "series"       "frequency"    "call"         "asy.var.coef"

> plot(m1$resid,type='l')  % Plot residuals of the fitted model (not shown)

> Box.test(m1$resid,lag=10,type='Ljung') % Model checking

        Box-Ljung test
data:  m1$resid
X-squared = 7.0808, df = 10, p-value = 0.7178

> m2=arima(x,order=c(3,0,0))  % Another approach with order given.
> m2
Call:  arima(x = x, order = c(3, 0, 0))

Coefficients:
```

```
        ar1     ar2      ar3  intercept  % Fitted model is
     0.3480  0.1793  -0.1423     0.0077  % y(t)=0.348y(t-1)+0.179y(t-2)
s.e.  0.0745  0.0778   0.0745     0.0012  %      -0.142y(t-3)+a(t),
                                         % where y(t) = x(t)-0.0077

sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,  aic = -1121.68
> names(m2)
 [1] "coef"      "sigma2"    "var.coef"  "mask"      "loglik"    "aic"
 [7] "arma"      "residuals" "call"      "series"    "code"      "n.cond"
[13] "model"

> Box.test(m2$residuals,lag=10,type='Ljung')
        Box-Ljung test
data:  m2$residuals
X-squared = 7.0169, df = 10, p-value = 0.7239

> ts.plot(m2$residuals)  % Residual plot

> tsdiag(m2) % obtain 3 plots of model checking (not shown in handout).

> p1=c(1,-m2$coef[1:3])  % Further analysis of the fitted model.
> roots=polyroot(p1)
> roots
[1]  1.590253+1.063882e+00i -1.920152-3.530887e-17i  1.590253-1.063882e+00i
> Mod(roots)
[1] 1.913308 1.920152 1.913308

> k=2*pi/acos(1.590253/1.913308)
> k
[1] 10.65638

> predict(m2,8)  % Prediction 1-step to 8-step ahead.
$pred
Time Series:
Start = 177
End = 184
Frequency = 1
[1] 0.001236254 0.004555519 0.007454906 0.007958518
[5] 0.008181442 0.007936845 0.007820046 0.007703826

$se
Time Series:
Start = 177
End = 184
Frequency = 1
[1] 0.009709322 0.010280510 0.010686305 0.010688994
```

```
[5] 0.010689733 0.010694771 0.010695511 0.010696190
```

Another example: Monthly U.S. unemployment rate from January 1948 to February, 2017. I use this example to emphasize two messages: (1) Modeling and prediction using AR models, including model simplification; (2) handling outliers.

Demonstration:

```
> require(quantmod)
> get Symbols("UNRATE",src="FRED")
> chartSeries(UNRATE)
> unrate <- as.numeric(UNRATE) ## create a regular vector, instead of a ‘‘xts’’ object
> tail(UNRATE)
           UNRATE
2016-09-01    4.9
2016-10-01    4.8
2016-11-01    4.6
2016-12-01    4.7
2017-01-01    4.8
2017-02-01    4.7
> tdx <- c(1:830)/12+1948
> plot(tdx,unrate,type='l',xlab='year',ylab='rate')
> title(main="Monthly U.S. unemployment rate")
> ar(unrate,method="mle")
Call:ar(x = unrate, method = "mle")
Coefficients:
      1        2        3        4        5        6        7        8
 0.9946   0.2152  -0.0713  -0.0533   0.0494  -0.1275  -0.0610   0.0513
      9       10       11
-0.0077  -0.1048   0.1003

Order selected 11  sigma^2 estimated as  0.03719
> m1 <- arima(unrate,order=c(11,0,0))
> m1
Call:arima(x = unrate, order = c(11, 0, 0))

Coefficients:
         ar1     ar2      ar3      ar4     ar5      ar6      ar7     ar8
      0.9945  0.2152  -0.0712  -0.0532  0.0493  -0.1275  -0.0610  0.0513
s.e.  0.0346  0.0488   0.0495   0.0495  0.0497   0.0495   0.0496  0.0496
          ar9     ar10    ar11  intercept
```

```
      -0.0077  -0.1047  0.1004     5.6715
s.e.   0.0496   0.0490  0.0348     0.4417

sigma^2 estimated as 0.03718:  log likelihood = 186.03,  aic = -346.07
> names(m1)
 [1] "coef"      "sigma2"    "var.coef"  "mask"      "loglik"    "aic"
 [7] "arma"      "residuals" "call"      "series"    "code"      "n.cond"
[13] "nobs"      "model"
> tsdiag(m1,gof=24)
> c1 <- c(NA,NA,NA,0,0,NA,0,0,0,NA,NA,NA)
> m2 <- arima(unrate,order=c(11,0,0),fixed=c1) ## refinement
> m2
Call:arima(x = unrate, order = c(11, 0, 0), fixed = c1)

Coefficients:
         ar1     ar2      ar3  ar4  ar5      ar6  ar7  ar8  ar9     ar10
      0.9967  0.2045  -0.0800    0    0  -0.1369    0    0    0  -0.0989
s.e.  0.0343  0.0481   0.0427    0    0   0.0291    0    0    0   0.0407
        ar11  intercept
      0.0998     5.6702
s.e.  0.0342     0.4416

sigma^2 estimated as 0.03733:  log likelihood = 184.37,  aic = -352.74
> tsdiag(m2)
> tsdiag(m2,gof=24)
> pm2 <- predict(m2,4)
> names(pm2)
[1] "pred" "se"
> low <- pm2$pred-1.96*pm2$se
> upp <- pm2$pred+1.96*pm2$se
> names(pm2)
[1] "pred" "se"
> pm2$pred
Time Series:
Start = 831
End = 834
Frequency = 1
[1] 4.737312 4.710012 4.745765 4.759146
> pm2$se
Time Series:
Start = 831
End = 834
Frequency = 1
[1] 0.1932128 0.2727943 0.3577585 0.4391267
> low
Time Series:
```

```
Start = 831
End = 834
Frequency = 1
[1] 4.358614 4.175335 4.044559 3.898457
> upp
Time Series:
Start = 831
End = 834
Frequency = 1
[1] 5.116009 5.244688 5.446972 5.619834
######################## Handling outliers
> which.min(m2$residuals)  ### locate the minimum of residuals
[1] 23
> I23 <- rep(0,830)
> I23[23] <- 1
> c1 <- c(c1,NA)
> m3 <- arima(unrate,order=c(11,0,0),fixed=c1,xreg=I23)
> m3
Call: arima(x = unrate, order = c(11, 0, 0), xreg = I23, fixed = c1)

Coefficients:
         ar1     ar2      ar3  ar4  ar5      ar6  ar7  ar8  ar9     ar10
      1.0449  0.1219  -0.0472    0    0  -0.1345    0    0    0  -0.1021
s.e.  0.0349  0.0515   0.0434    0    0   0.0287    0    0    0   0.0413
        ar11  intercept      I23
      0.1025     5.6709  -0.7749
s.e.  0.0345     0.4428   0.1338

sigma^2 estimated as 0.03591:  log likelihood = 200.43,  aic = -382.87
> tsdiag(m3,gof=24)
> which.max(m3$residuals) ### locate the maximum of the residuals
[1] 22
> I22 <- rep(0,830)
> I22[22] <- 1
> c1 <- c(c1,NA)
> X <- cbind(I23,I22)
> m4 <- arima(unrate,order=c(11,0,0),fixed=c1,xreg=X)
> m4
Call:arima(x = unrate, order = c(11, 0, 0), xreg = X, fixed = c1)

Coefficients:
         ar1     ar2      ar3  ar4  ar5      ar6  ar7  ar8  ar9     ar10
      1.0764  0.1170  -0.0955    0    0  -0.1069    0    0    0  -0.0951
s.e.  0.0346  0.0507   0.0431    0    0   0.0283    0    0    0   0.0416
        ar11  intercept      I23     I22
      0.0901     5.6690  -0.2580  1.1729
```

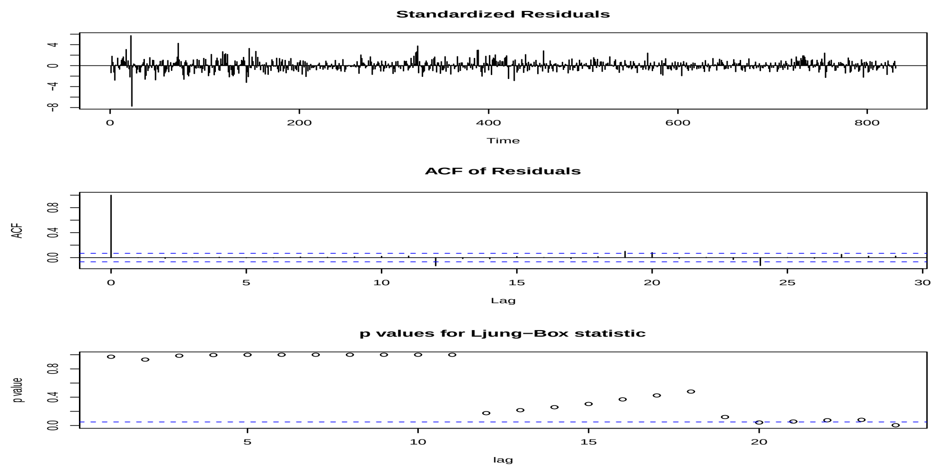


Figure 4: Model checking for AR(11) model fitted to UNRATE series.

```
s.e.  0.0348     0.4388   0.1367  0.1375

sigma^2 estimated as 0.03305:  log likelihood = 234.86,  aic = -449.73
> tsdiag(m4,gof=24)
```

Moving-average (MA) model

Model with finite memory!

Some daily stock returns have minor serial correlations and can be modeled as MA or AR models.

MA(1) model

- Form: $r_t = \mu + a_t - \theta a_{t-1}$
- Stationarity: always stationary.
- Mean (or expectation): $E(r_t) = \mu$
- Variance: $\text{Var}(r_t) = (1 + \theta^2)\sigma_a^2$.
- Autocovariance:

1. Lag 1: $\text{Cov}(r_t, r_{t-1}) = -\theta\sigma_a^2$
2. Lag ℓ: $\text{Cov}(r_t, r_{t-\ell}) = 0$ for $\ell > 1$.

Thus, r_t is not related to $r_{t-2}, r_{t-3}, \cdots$.

- ACF: $\rho_1 = \frac{-\theta}{1+\theta^2}$, $\rho_\ell = 0$ for $\ell > 1$.

  Finite memory! MA(1) models do not remember what happen two time periods ago.

- Forecast (at origin $t = n$):

  1. 1-step ahead: $\hat{r}_n(1) = \mu - \theta a_n$. Why? Because at time n, a_n is known, but a_{n+1} is not.
  2. 1-step ahead forecast error: $e_n(1) = a_{n+1}$ with variance σ_a^2.
  3. Multi-step ahead: $\hat{r}_n(\ell) = \mu$ for $\ell > 1$.

     Thus, for an MA(1) model, the multi-step ahead forecasts are just the mean of the series. Why? Because the model has memory of 1 time period.
  4. Multi-step ahead forecast error:

     $$e_n(\ell) = a_{n+\ell} - \theta a_{n+\ell-1}$$
  5. Variance of multi-step ahead forecast error:

     $(1 + \theta^2)\sigma_a^2 =$ variance of r_t.

- Invertibility:

  - Concept: r_t is a proper linear combination of a_t and the past observations $\{r_{t-1}, r_{t-2}, \cdots\}$.

– Why is it important? It provides a simple way to obtain the shock a_t.

For an invertible model, the dependence of r_t on $r_{t-\ell}$ converges to zero as ℓ increases.

– Condition: $|\theta| < 1$.

– Invertibility of MA models is the dual property of stationarity for AR models.

MA(2) model

- Form: $r_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$. or

$$r_t = \mu + (1 - \theta_1 B - \theta_2 B^2) a_t.$$

- Stationary with $E(r_t) = \mu$.
- Variance: $\text{Var}(r_t) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$.
- ACF: $\rho_2 \neq 0$,but $\rho_\ell = 0$ for $\ell > 2$.
- Forecasts go the the mean after 2 periods.

Building an MA model

- Specification: Use sample ACF

  Sample ACFs are all small after lag q for an MA(q) series. (See test of ACF.)

- Constant term? Check the sample mean.

- Estimation: use maximum likelihood method
  - Conditional: Assume $a_t = 0$ for $t \le 0$
  - Exact: Treat a_t with $t \le 0$ as parameters, estimate them to obtain the likelihood function.

  Exact method is preferred, but it is more computing intensive.
- Model checking: examine residuals (to be white noise)
- Forecast: use the residuals as $\{a_t\}$ (which can be obtained from the data and fitted parameters) to perform forecasts.

Model form in R: R parameterizes the MA(q) model as

$$r_t = \mu + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$

instead of the usual *minus* sign in θ. Consequently, care needs to be exercised in writing down a fitted MA parameter in R. For instance, an estimate $\hat{\theta}_1 = -0.5$ of an MA(1) in R indicates the model is $r_t = a_t - 0.5a_{t-1}$.

Example:Daily log return of the value-weighted index
R demonstration

```
> setwd("your working directory")
> library(fBasics)
> da=read.table("d-ibmvwew6202.txt")
> dim(da)
[1] 10194     4

> vw=log(1+da[,3])*100  % Compute percentage log returns of the vw index.
> acf(vw,lag.max=10)  % ACF plot is not shon in this handout.
> m1=arima(vw,order=c(0,0,1))  % fits an MA(1) model
> m1
```

```
Call:
arima(x = vw, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.1465     0.0396 % The model is vw(t) = 0.0396+a(t)+0.1465a(t-1).
s.e.  0.0099     0.0100

sigma^2 estimated as 0.7785:  log likelihood = -13188.48,  aic = 26382.96
> tsdiag(m1)
> predict(m1,5)
$pred
Time Series:
Start = 10195
End = 10199
Frequency = 1
[1] 0.05036298 0.03960887 0.03960887 0.03960887 0.03960887

$se
Time Series:
Start = 10195
End = 10199
Frequency = 1
[1] 0.8823290 0.8917523 0.8917523 0.8917523 0.8917523
```

Mixed ARMA model: A compact form for flexible models. Focus on the ARMA(1,1) model for

1. simplicity
2. useful for understanding GARCH models in Ch. 3 for volatility modeling.

<u>ARMA(1,1) model</u>

- Form: $(1-\phi_1 B)r_t = \phi_0 + (1-\theta B)a_t$ or

$$r_t = \phi_1 r_{t-1} + \phi_0 + a_t - \theta_1 a_{t-1}.$$

A combination of an AR(1) on the LHS and an MA(1) on the RHS.

- Stationarity: same as AR(1)
- Invertibility: same as MA(1)
- Mean: as AR(1), i.e. $E(r_t) = \frac{\phi_0}{1-\phi_1}$
- Variance: given in the text
- ACF: Satisfies $\rho_k = \phi_1 \rho_{k-1}$ for $k > 1$, but

$$\rho_1 = \phi_1 - [\theta_1 \sigma_a^2 / \mathrm{Var}(r_t)] \neq \phi_1.$$

  This is the difference between AR(1) and ARMA(1,1) models.
- PACF: does not cut off at finite lags.

Building an ARMA(1,1) model

- Specification: use EACF or AIC
- Use the command **auto.arima** of the package **forecast**.
- Estimation: cond. or exact likelihood method
- Model checking: as before
- Forecast: MA(1) affects the 1-step ahead forecast. Others are similar to those of AR(1) models.

Three model representations:

- ARMA form: compact, useful in estimation and forecasting
- AR representation: (by long division)

$$r_t = \phi_0 + a_t + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \cdots$$

  It tells how r_t depends on its past values.

- MA representation: (by long division)

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

  It tells how r_t depends on the past shocks.

For a stationary series, ψ_i converges to zero as $i \to \infty$. Thus, the effect of any shock is transitory.
The MA representation is particularly useful in computing variances of forecast errors.
For a ℓ-step ahead forecast, the forecast error is

$$e_n(\ell) = a_{n+\ell} + \psi_1 a_{n+\ell-1} + \cdots + \psi_{\ell-1} a_{n+1}.$$

The variance of forecast error is

$$\mathrm{Var}[e_n(\ell)] = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma_a^2.$$

Unit-root Nonstationarity

<u>Random walk</u>

- Form $p_t = p_{t-1} + a_t$
- Unit root? It is an AR(1) model with coefficient $\phi_1 = 1$.

- Nonstationary: Why? Because the variance of r_t diverges to infinity as t increases.
- Strong memory: sample ACF approaches 1 for any finite lag.
- Repeated substitution shows

$$p_t = \sum_{i=0}^{\infty} a_{t-i} = \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

  where $\psi_i = 1$ for all i. Thus, ψ_i does not converge to zero. The effect of any shock is permanent.

<u>Random walk with drift</u>

- Form: $p_t = \mu + p_{t-1} + a_t, \quad \mu \neq 0$.
- Has a unit root
- Nonstationary
- Strong memory
- Has a time trend with slope μ. Why?

<u>differencing</u>

- 1st difference: $r_t = p_t - p_{t-1}$

  If p_t is the log price, then the 1st difference is simply the log return. Typically, 1st difference means the "change" or "increment" of the original series.

- Seasonal difference: $y_t = p_t - p_{t-s}$, where s is the periodicity, e.g. $s = 4$ for quarterly series and $s = 12$ for monthly series.

  If p_t denotes quarterly earnings, then y_t is the change in earning from the same quarter one year before.

Meaning of the constant term in a model

- MA model: mean
- AR model: related to mean
- 1st differenced: time slope, etc.

Practical implication in financial time series
Example: Monthly log returns of General Electrics (GE) from 1926 to 1999 (74 years)
Sample mean: 1.04%, std($\hat{\mu}$) = 0.26
Very significant!
is about 12.45% a year
$1 investment in the beginning of 1926 is worth

- annual compounded payment: $5907
- quarterly compounded payment: $8720
- monthly compounded payment: $9570
- Continuously compounded?

Unit-root test

Let p_t be the log price of an asset. To test that p_t is not predictable (i.e. has a unit root), two models are commonly employed:

$$\begin{aligned} p_t &= \phi_1 p_{t-1} + e_t \\ p_t &= \phi_0 + \phi_1 p_{t-1} + e_t. \end{aligned}$$

The hypothesis of interest is $H_o : \phi_1 = 1$ vs $H_a : \phi_1 < 1$.

Dickey-Fuller test is the usual t-ratio of the OLS estimate of ϕ_1 being 1. This is the DF unit-root test. The t-ratio, however, has a non-standard limiting distribution.

Let $\Delta p_t = p_t - p_{t-1}$. Then, the augmented DF unit-root test for an AR(p) model is based on

$$\Delta p_t = c_t + \beta p_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta p_{t-i} + e_t.$$

The t-ratio of the OLS estimate of β is the ADF unit-root test statistic. Again, the statistic has a non-standard limiting distribution.

Example: Consider the log series of U.S. quaterly real GDP series from 1947.I to 2009.IV. (data from Federal Reserve Bank of St. Louis). See `q-gdpc96.txt` on the course web.

R demonstration

```
> library(fUnitRoots)
> help(UnitrootTests)  % See the tests available
>da=read.table(‘‘q-gdpc96.txt’’,header=T)
>gdp=log(da[,4])

> adfTest(gdp,lag=4,type=c("c")) #Assume an AR(4) model for the series.

Title: Augmented Dickey-Fuller Test
```

```
Test Results:
  PARAMETER:
    Lag Order: 4
  STATISTIC:
    Dickey-Fuller: -1.7433
  P VALUE:
    0.4076  # cannot reject the null hypothesis of a unit root.

*** A more careful analysis
> x=diff(gdp)
> ord=ar(x)   # identify an AR model for the differenced series.
> ord

Call:ar(x = x)

Coefficients:
      1        2        3
 0.3429   0.1238  -0.1226

Order selected 3  sigma^2 estimated as  8.522e-05
>   # An AR(3) for the differenced data is confirmed.
    # Our previous analysis is justified.
```

Discussion: The command `arima` on R.

1. Dealing with the constant term. If there is any differencing, no constant is used.

   The subcommand `include.mean=T` in the `arima` command.

2. Fixing some parameters. Use subcommand `fixed` in `arima`.

   See the unemployment rate series used in AR modeling.