# Time Series Analysis

Lecture 8

# Review

1. Autoregression Conditional Heteroskedasticity(ARCH)
2. Extensions(GARCH,IARCH,ARCH-M,EGARCH)

# Today's Topics

1. Long Memory Models
2. Seasonal Models

# Today's Topics

# Long Memory Processes: Motivation

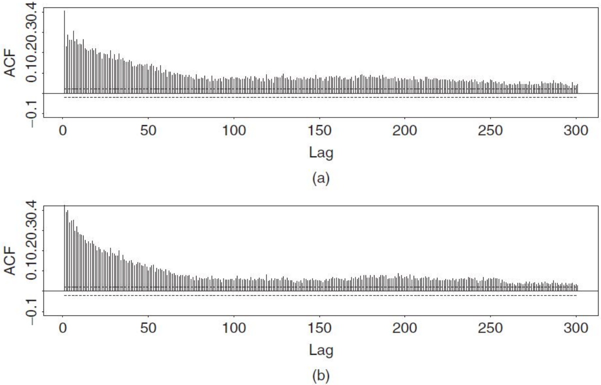


**Figure 2.22** Sample autocorrelation function of absolute series of daily simple returns for CRSP value- and equal-weighted indexes: (a) value-weighted index return and (b) equal-weighted index return. Sample period is from January 2, 1970, to December 31, 2008.

1. Figure 2.22 shows the sample ACFs of the absolute series of daily simple returns for the CRSP value- and equal-weighted indexes from January 2, 1970, to December 31, 2008.
2. The ACFs are relatively small in magnitude but decay very slowly; they appear to be significant at the 5% level even after 300 lags.

# Today's Topics

1. Long Memory Models
    - Motivation
    - Definition of long memory
    - Fractional Integration
    - Properties of Fractional Integrated Series
    - Estimation of $d$
    - Fractional Brownian Motion
    - Fractional Dicky-Fuller Test
    - The Augmented FD-F Test
2. Seasonal Models

# Definition of long memory

Given a discrete time series process $y_t$ with autocorrelation function $\rho_j$ at lag $j$, then according to McLeod and Hipel (1978), the process possesses long memory if the quantity

$$\lim_{n\to\infty} \sum_{j=-n}^{n} |\rho_j|,$$

is nonfinite.

A stationary and invertible ARMA process has autocorrelations which are geometrically bounded, i.e., $|\rho_k| \leq cm^{-k}$, for large $k$, where $0 < m < 1$ and is hence a short memory process.

# Today's Topics

# Fractional Integration: $I(d)$

A process $x_t$ is said to be integrated of order $d$, or $I(d)$, if

$$(1-B)^d x_t = a_t, \tag{1}$$

where $a_t$ is a stationary and ergodic process with a bounded and positively valued spectrum at all frequencies (Granger and Joyeux, 1980; Hosking, 1981).

1. For $0 < d < 0.5$, the process is long memory, its autocorrelations are all positive and decay at a hyperbolic rate.
2. For $-0.5 < d < 0$, the sum of absolute values of the processes autocorrelations tends to a constant, so that it has short memory.

Consider the $MA(\infty)$ representation implied by (1). The inverse of the operator $(1-B)^d$ exists provided that $d < 1/2$. (1) can be rewritten as

$$x_t = (1-B)^{-d} a_t.$$

For $z$ a scalar, define the function

$$f(z) \equiv (1-z)^{-d}.$$

This function has derivatives given by

$$\begin{aligned}
\frac{\partial f}{\partial z} &= d \cdot (1-z)^{-d-1}, \\
\frac{\partial^2 f}{\partial z^2} &= (d+1) \cdot d \cdot (1-z)^{-d-2}, \\
\frac{\partial^3 f}{\partial z^3} &= (d+2) \cdot (d+1) \cdot d \cdot (1-z)^{-d-3}, \\
&\vdots \\
\frac{\partial^j f}{\partial z^j} &= (d+j-1) \cdot (d+j-2) \cdots (d+1) \cdot d \cdot (1-z)^{-d-j}.
\end{aligned}$$

A power series expansion for $f(z)$ around $z=0$ is thus given by

$$\begin{aligned}(1-z)^{-d} &= f(0)+\left.\frac{\partial f}{\partial z}\right|_{z=0}\cdot z+\frac{1}{2!}\left.\frac{\partial^2 f}{\partial z^2}\right|_{z=0}\cdot z^2+\frac{1}{3!}\left.\frac{\partial^3 f}{\partial z^3}\right|_{z=0}\cdot z^3+\cdots\\ &= 1+d\cdot z+(1/2!)(d+1)d\cdot z^2+(1/3!)(d+2)(d+1)d\cdot z^3+\cdot\end{aligned}$$

This suggests that the operator $(1-B)^{-d}$ might be represented by the filter

$$\begin{aligned}(1-B)^{-d} &= 1+dB+(1/2!)(d+1)dB^2\\ &\quad +(1/3!)(d+2)(d+1)dB^3+\cdots\\ &= \sum_{j=0}^{\infty} h_j B^j,\end{aligned}$$

where $h_0 \equiv 1$ and

$$h_j \equiv (1/j!)(d+j-1)(d+j-2)(d+j-3)\cdots(d+1)(d).$$

Note that if $d < 1$, $h_j$ can be approximated for large $j$ by

$$h_j \cong (j+1)^{d-1}.$$

Thus, the time series model

$$x_t = (1-B)^{-d} a_t = h_0 \varepsilon_t + h_1 \varepsilon_{t-1} + h_2 \varepsilon_{t-2} + \cdots ,$$

describes an $MA(\infty)$ representation in which the impulse-response coefficient $h_j$ behave for large $j$ like $(j+1)^{d-1}$. Granger and Joyeux proposed the fractionally integrated process as an approach to modeling long memories in a time series.

Recall that the impulse-response coefficient associated with the $AR(1)$ process $x_t = (1-\phi B)^{-1} a_t$ is given by $\phi^j$. The impulse-response coefficients for a stationary $ARMA$ process decay geometrically, in contrast to the slower decay implied by fractional integration process.

In a finite sample, this long memory could be approximated arbitrarily well with a suitable large-order *ARMA* representation. The goal of the fractional-difference specification is to capture parsimoniously long-run multipliers that decay very slowly. Square-summable provided that $d < 1/2$:

$$\sum_{j=0}^{\infty} h_j^2 < \infty \qquad \textit{for } d < 1/2.$$

Thus, defines a covariance-stationary process provided that $d < 1/2$. If $d > 1/2$, the proposal is to difference the process before describing it. For example, if $d = 0.7$, the process implies

$$(1-B)^{-0.3}(1-B)x_t = \psi(B)a_t;$$

that is, $\nabla x_t$ is fractionally integrated with parameter $d = -0.3 < 1/2$.

# Today's Topics

1. Long Memory Models
   - Motivation
   - Definition of long memory
   - Fractional Integration
   - Properties of Fractional Integrated Series
   - Estimation of $d$
   - Fractional Brownian Motion
   - Fractional Dicky-Fuller Test
   - The Augmented FD-F Test

2. Seasonal Models

## Properties of $I(d)$ $(-\frac{1}{2} < d < \frac{1}{2})$

$$(1-B)^d x_t = a_t, \tag{2}$$

where $a_t$ is a white noise process consist of i.i.d distributed rvs with mean zero and finite variance $\sigma^2$. For convenience, assume that $\sigma^2 = 1$.

(a) When $d < 1/2$, $\{x_t\}$ is a stationary process and has the infinite moving-average representation

$$x_t = \psi(B)a_t = \sum_{k=0}^{\infty} \psi_k a_{t-k},$$

where

$$\psi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}$$

As $k \to \infty$, $\psi_k \sim k^{d-1}/(d-1)!$.

(b) When $d > -1/2$, $\{x_t\}$ is invertible and has the infinite autoregressive representation

$$\pi(B)x_t = \sum_{k=1}^{\infty} \pi_k x_{t-k} = a_t,$$

where

$$\pi_k = \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}.$$

As $k \to \infty$. $\pi_k \sim k_{-d-1}/(-d-1)!$. In parts (c)-(f), we assume that $-\frac{1}{2} < d < \frac{1}{2}$.

(c) The covariance function of $\{x_t\}$ is

$$\gamma_k = E(x_t x_{t-k}) = \frac{(-1)^k(-2d)!}{(k-d)!(-k-d)!},$$

and the correlation function of $\{x_t\}$ is

$$\begin{aligned} \rho_k &= \gamma_k/\gamma_0 = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!} \quad (k = 0, \pm 1, \cdots), \\ \rho_k &= \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)} \quad (k = 1, 2, \cdots). \end{aligned}$$

In particular $\gamma_0 = (-2d)!/\{(-d)!\}^2$ and $\rho_1 = d/(1-d)$. As $k \to \infty$,

$$\rho_k \sim \frac{(-d)!}{(d-1)!} k^{2d-1}.$$

(d) The partial correlations of $\{x_t\}$ are

$$\phi_{kk} = d/(k-d) \quad (k = 1, 2, \cdots).$$

# Properties of $I(d)$ $(d = -1/2)$

Given $(1-B)^{-1/2}x_t = a_t$, where $a_t$ is i.i.d rv with finite variance $\sigma^2$, we have

(a) $x_t$ is stationary and has an MA representation as $\sum_{j=-\infty}^{\infty} c_{t-j}a_j$ where

$$c_j = \frac{\Gamma(j-1/2)}{j!\Gamma(-1/2)} \qquad \textit{and} \qquad c_j \sim \frac{j^{-3/2}}{\Gamma(-1/2)},$$

as $j \to \infty$.

(b) The covariance structure of $x_t$ is

$$\sigma_k = E(x_t x_{t-k}) = \frac{4\sigma^2}{\pi}\frac{1}{1-4k^2}.$$

# Today's Topics

# Estimation of $d$ (i.i.d case)

$$(1-B)^d x_j = a_j, \quad (j = 1, 2, \cdots, T),$$

where $d$ is any value and $\{a_j\} \sim NID(0, \sigma^2)$. The parameter estimated here are $d$ and $\sigma^2$, and the concentrated log-likelihood for $d$ is given, except for constants, by

$$\ell(d) = -\frac{T}{2} \log \left[ \sum_{j=1}^{T} \{(1-B)^d x_j\}^2 \right]$$

Theorem: Let $\hat{d}$ be the MLE of $d_0$. Then it holds that, as $T \to \infty$,

$$\sqrt{T}(\hat{d} - d_0) \to N(0, 6/\pi^2).$$

# Estimation of $d$ (Stationary case)

$$(1-B)^d a(B) x_j = b(B) a_j, \quad (j = 1, 2, \cdots, T),$$

where $a_j \sim NID(0, \sigma^2)$, $a(B) = 1 - a_1 B - \cdots - a_p B^p$, and $b(B) = 1 - b_1 B - \cdots - b_q B^q$ with $a(x) \neq 0$ and $b(x) \neq 0$ for $|x| < 1$. The parameters to be estimated are $d, \psi = (a_1, \cdots, a_p, b_1, \cdots, b_q)'$, and $\sigma^2$. The concentrated log-likelihood for $d$ and $\psi$ is now given, except for constants, by

$$\ell(d, \psi) = -\frac{T}{2} \log \left[ \sum_{j=1}^{T} \{a(B) b^{-1}(B)(1-B)^d x_j\}^2 \right].$$

Theorem: The MLE's of $d_0$ and $\psi_0$ are asymptotically unique,and let $\hat{\tau} = (\hat{d}, \hat{\psi}')'$ be the MLE of $\tau_0 = (d_0, \psi_0')'$. Then is holds that, as $T \to \infty$,

$$\sqrt{T}(\hat{\tau} - \tau_0) \to N(0, \Omega^{-1}),$$

where

$$\Omega = \begin{bmatrix} \pi^2/6 & \kappa' \\ \kappa & \Phi \end{bmatrix}$$

with $\kappa = (\kappa_1, \cdots, \kappa_p, \lambda_1, \cdots, \lambda_q)'$,
$\kappa_i = \sum_{j=1}^{\infty} \frac{1}{j} c_{j-i}, \qquad \lambda_i = -\sum_{j=i}^{\infty} \frac{1}{j} d_{j-i}$, with $c_j$ and $d_j$ the coefficients of $B^j$ in the expansion of $1/a(B)$ and $1/b(B)$, respectively, and $\Phi$ the Fisher information matrix for $a$ and $b$.

# Today's Topics

# Fractional Brownian Motion

In probability theory, fractional Brownian motion (fBm), also called a fractal Brownian motion, is a generalization of Brownian motion. Unlike classical Brownian motion,the increments of fBm need not be independent. fBm is a continuous-time Gaussian process $B_H(t)$ on $[0, T]$, which starts at zero, has expectation zero for all $t$ in $[0, T]$, and has the following covariance function:

$$E[B_H(t)B_H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}),$$

where $d$ is a real number in $(0, 1)$, called the Hurst index or Hurst parameter associated with the fractional Brownian motion. The Hurst exponent describes the raggedness of the resultant motion, with a higher value leading to a smoother motion. It was introduced by Mandelbrot & Van Ness (1968).

In a sense fractional Brownian motion, $B_H(r)$, can be regarded as the approximate $(1/2 - H)$ fractional derivative of regular Brownian motion,

$$B_H(r) = [1/\Gamma(H + 1/2)] \int_0^r (r - x)^{H-1/2} dB(x) \qquad \textit{for } r \in (0, 1),$$

where $\Gamma(\cdot)$ is the gamma function, $B(x)$ is regular Brownian motion with unit variance.
The autocovariance function of fractional Brownian motion is given by

$$E|B_H(t) - B_H(s)|^2 = |t - sl^{2H},$$

and

$$\gamma_k \approx |k|^{2H-2},$$

so that for high lags hyperbolic decay occurs in the autocovariance function.

The value of H determines what kind of process the *fBm* is:

(a) if $H = 1/2$, then the process is in fact a Brownian motion or Wiener process;

(b) if $H > 1/2$, then the increments of the process are positively correlated;

(c) if $H < 1/2$, then the increments of the process are negatively correlated.

The increment process, $X(t) = B_H(t+1) - B_H(t)$, is known as fractional Gaussian noise.

# Plots of fBM

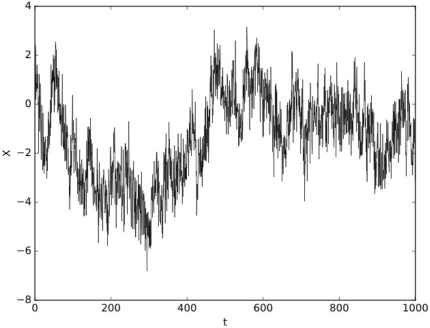


Figure 1 : $H = 0.15$

# Plots of fBM

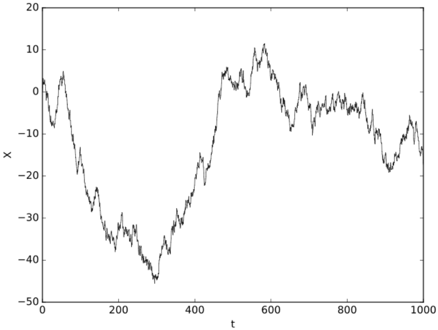


Figure 2 : $H = 0.55$

# Plots of fBM

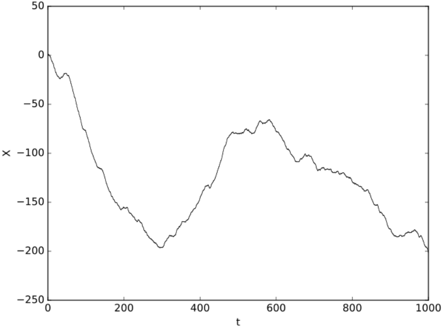


Figure 3 : $H = 0.95$

# Some asymptotic distributions

$$(1-B)^d x_t = a_t, \quad (1-B)y_t = x_t,$$

where $a_t$ is a white noise with zero mean and variance $\sigma^2$. Obviously, $y_t$ is nonstatioary.

And Sowell (1990) proves that

$$\sigma_y^2 \sim \sigma^2 \frac{\Gamma(1-2d)}{(1+2d)(\Gamma(1+d))(\Gamma(1-d))} T^{1+2d}, \quad \sigma_y^2 = Var(y_T).$$

Davydov (1970) shows that as $T \to \infty$,

$$\frac{1}{\sigma_y} y_{[Tr]} \Rightarrow B_d(r).$$

$B_d(r)$ is the normalized fractional Brownian motion which is defined by the following stochastic integral:

$$B_d(r) = \{1/\Gamma(d+1)\} \int_0^r (r-x)^d dB(x),$$

where the fractional differencing parameter $d$ is related to the Hurst coefficient as $d = H - 1/2$.

(a) $\sum_{t=1}^{T} \frac{x_t}{\sigma_y} \Rightarrow B_d(1)$

(b) $\frac{1}{T} \sum_{t=1}^{T} x_t^2 \xrightarrow{p} \gamma_x(0) = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} \sigma_a^2$

(c) $\frac{1}{T} \sum_{t=1}^{T} (x_t^2 - \bar{x})^2 \xrightarrow{p} \gamma_x(0)$

(d) $\sum_{t=1}^{T} \frac{t x_t}{T \sigma_y} \Rightarrow B_d(1) - \int_0^1 B_d(s) ds$

(e) $\frac{1}{T} \sum_{t=1}^{T} \frac{y_t}{\sigma_y} \Rightarrow \int_0^1 B_d(s) ds$

(f) $\frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\sigma_y^2} \Rightarrow \int_0^1 [B_d(s)]^2 ds$

(g) $\frac{1}{T} \sum_{t=1}^{T} \frac{(y_t - \bar{y})^2}{\sigma_y^2} \Rightarrow \int_0^1 [B_d(s)]^2 ds - \left[ \int_0^1 B_d(s) ds \right]^2$

# Today's Topics

# Fractional Dicky-Fuller Test

In its simplest formulation, the D-F statistic is based upon testing for the statistical significance of the parameter $\phi$ in the following regression model:

$$\Delta x_t = \phi x_{t-1} + a_t. \tag{3}$$

If $a_t$ is i.i.d., $x_t$ is a random walk when $\phi = 0$ whilst $x_t$ is a stationary $AR(1)$ process if $\phi < 0$.

In what follows, we generalize the regression model in (3) to test the null hypothesis that a series is $I(d_0)$ against the alternative that it is $I(d_1)$, with $d_1 < d_0$. Specifically, our proposal is based upon testing for the statistical significance of coefficient $\phi$ in the following regression:

$$\Delta^{d_0} x_t = \phi \Delta^{d_1} x_{t-1} + a_t. \tag{4}$$

where $a_t$ is an $I(0)$ process.

When $\phi = 0$, the series follows the process

$$\Delta^{d_0} x_t = a_t,$$

implying that $x_t$ is $I(d_0)$. When $\phi < 0$, $x_t$ can be expressed as

$$(\Delta^{d_0 - d_1} - \phi B)\Delta^{d_1} x_t = a_t,$$

The polynomial $\Pi(z) = ((1 - z)^{d_0 - d_1} - \phi z)$ has absolutely summable coefficients and verifies $\Pi(0) = 1$ and $\Pi(1) \neq 0$. All the roots of the polynomial are outside the unit circle if $-2^{d_0 - d_1} < \phi < 0$. As in the D-F framework, this condition excludes explosive processes.

Under the previous restriction on $\phi$, $\Delta^{d_1} x_t$, is $I(0)$, so that $x_t$ is an $I(d_1)$ process that can be rewritten as

$$\Delta^{d_1} x_t = C(B) a_t,$$

where $C(z) = \Pi(z)^{-1} = ((1 - z)^{d_0 - d_1} - \phi z)^{-1}$, with $C(0) = 1$ and $0 < C(1) < \infty$.

# Fractional Dicky-Fuller Test

DGP:

$$(1-B)^{d_0} x_t = a_t,$$

Tesing Model:

$$\Delta^{d_0} x_t = \phi \Delta^{d_1} x_{t-1} + a_t, \tag{5}$$

where $\{a_t\}$ is a sequence of zero-mean i.i.d. random variables with unknown variance $\sigma^2$ and the finite fourth-order moment.

$$H_0 : \ \phi = 0, \ x_t \ is \ I(d_0),$$
$$H_1 : \ \phi < 0, \ x_t \ is \ I(d_1),$$

such that when $d_0 = 1$ and $d_1 = 0$ the conventional $I(1)$ vs. $I(0)$ framework is recovered.

Let us consider $d_0 = 1$.The OLS estimator of $\phi, \hat{\phi}_{ols}$ and its $t$-ratio, $t_{\hat{\phi}_{ols}}$, are given by the usual least-squares expressions

$$
\begin{aligned}
\hat{\phi}_{ols} &= [\sum_{t=2}^{T} \Delta x_t \Delta^{d_1} x_{t-1}]/[\sum_{t=2}^{T} (\Delta^{d_1} x_{t-1})^2], \\
\hat{t}_{\phi_{ols}} &= [\sum_{t=2}^{T} \Delta x_t \Delta^{d_1} x_{t-1}]/[S_T \Big( \sum_{t=2}^{T} (\Delta^{d_1} x_{t-1})^2 \Big)^{1/2}],
\end{aligned}
$$

where the variance of the residuals, $S_T^2$, is given by

$$
S_T^2 = [\sum (\Delta x_t - \hat{\phi}_{ols} \Delta^{d_1} x_{t-1})^2]/T.
$$

Theorem : Under the null hypothesis that $y_t$ is a random walk ($d_0 = 1$), $\hat{\phi}_{ols}$ is a consistent estimator of $\phi = 0$ and converges to its true value at a rate $T^{1-d_1}$ when $0 < d_1 < 0.5$, $(T \log T)^{1/2}$ when $d_1 = 0.5$, and at the standard rate $T^{1/2}$ when $0.5 < d_1 < 1$. Its asymptotic distribution is given by

$$\begin{aligned} T^{1-d_1}\hat{\phi}_{ols} &\Rightarrow \int_0^1 B_{-d_1}(r)dB(r), && \textit{if } 0 \leq d_1 \leq 0.5, \\ (T \log T)^{1/2}\hat{\phi}_{ols} &\Rightarrow N(0, \pi), && \textit{if } d_1 = 0.5, \\ T^{1/2}\hat{\phi}_{ols} &\Rightarrow N(0, \frac{\Gamma^2(d_1)}{\Gamma(2d_1 - 1)}), && \textit{if } 0.5 \leq d_1 \leq 1. \end{aligned}$$

Under the null hypothesis that $y_t$ is a random walk ($I(1)$, i.e., $d_0 = 1$), the asymptotic distribution of $t_{\hat{\phi}_{ols}}$ is given by

$$
\begin{aligned}
t_{\hat{\phi}_{ols}} &\Rightarrow \frac{\int_0^1 B_{-d_1}(r)dB(r)}{(\int_0^1 B_{-d_1}^2(r)dr)^{1/2}}, \quad \mathit{if}\ 0 \leq d_1 < 0.5, \\
t_{\hat{\phi}_{ols}} &\Rightarrow N(0,1) \qquad \mathit{if}\ 0.5 \leq d_1 < 1.
\end{aligned}
$$

| | | FD-F | | | | |
|---|---|---|---|---|---|---|
| $T = 100$ | $d_1$ | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| | size | 5.6% | 5.9% | 5.5% | 4.9% | 5.3% |
| $T = 400$ | $d_1$ | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| | size | 5.6% | 5.2% | 5.1% | 5.0% | 5.4% |

Figure 4 : Size of Unit Root Tests against Fractional Alternatives (Dolado et. al., 2002, Econometrica)

# The FD-F with unknown $d_1$

To implement the FD-F test, we need a value of the memory parameter $d_1$ under the alternative hypothesis. Sometimes, we can have a priori knowledge of this value but, more realistically, we will need to estimate it, particularly when a composite alternative is being posed.

# Asymptotic Distribution of the FD-F Test with an Estimated $d_1$

Let us assume that the DGP is a random walk such that $\Delta x_t = a_t$. Let $\hat{d}_T$ be a $T^{1/2}$-consistent estimator of $d_1 < 1$ (e.g. the MLE). Since the value of $d_1$ that is needed to implement the test ought to be strictly smaller than 1, the pre-estimated value of $d_1$, $\hat{d}_1$, is chosen in accord with the following trimming rule:

$$\hat{d}_1 = \begin{cases} \hat{d}_T, \quad if \ \hat{d}_T < 1 - c & ; \\ 1 - c, if \ \hat{d}_T \geq 1 - c & . \end{cases}$$

where $c > 0$ is a (fixed) value in a neighborhood of zero such that $(1 - c)$ is sufficiently close to unity.

Theorem: Under the null hypothesis that $x_t$ is a random walk, the test statistic $t_{\hat{\phi}_{ols}}$ associated to parameter $\phi$ in the regression

$$\Delta x_t = \phi \Delta^{\hat{d}_1} x_{t-1} + a_t,$$

where $\hat{d}_1$ has been chosen according to the above criterion, is asymptotically distributed as

$$t_{\hat{\phi}_{ols}} \Rightarrow N(0,1).$$

As a result, the standard normal tables are applicable.

# Today's Topics

# The Augmented FD-F Test

In several applications, such as in the case of financial or opinion poll data, the characterization of a series as a fractional white noise may well be plausible. However, in many other cases, it seems of paramount importance to allow for more general models where there is serial correlation in the errors.

# ADF Test: Review

Let us consider again regression:

$$\Delta x_t = \phi x_{t-1} + u_t,$$

where now $\alpha(B)u_t = a_t$, i.e., an autoregressive process of order $p$, $AR(p)$, such that $\alpha(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^P$ has all its roots outside the unit circle. Then the AD-F test is based upon the following regression:

$$\Delta x_t = \phi x_{t-1} + \sum_{i=1}^{p} \zeta_i \Delta x_{t-i} + a_t.$$

Under the null hypothesis, Dickey and Fuller (1981) showed that the asymptotic distribution of the $t$-ratio $t_{\hat{\phi}_{ols}}$ is identical to the one derived in the absence of serial correlation.

Following the same analogy that was used in FD-F test, the proposed regression for the case where $u_t$ is an $AR(p)$ process is as follows:

$$\Delta x_t = \phi \Delta^{d_1} x_{t-1} + \sum_{i=1}^{p} \zeta_i \Delta x_{t-i} + a_t. \qquad (6)$$

When $\phi = 0$, it is easy to prove that $x_t$ follows $ARIMA(p, 1, 0)$ process, whereas, when $\phi < 0$, $x_t$ is given by

$$(\alpha(B)\Delta^{1-d_1} - \phi B)\Delta^{d_1} x_t = u_t,$$

where, as in FD-F test, the polynomial $\Pi(z) = (\alpha(z)(1-z)^{1-d_1} - \phi z)$ has absolutely summable coefficients which verify that $\Pi(0) = 1, \Pi(1) = -\phi$, and is an invertible process that, under $H_1$, can be rewritten as

$$\Delta^{d_1} x_t = C(B) a_t,$$

where, in this case, $C(z) = (\Pi(z))^{-1}\alpha(z)$ such that $C(0) = 1$ and $0 < C(1) < 0$.

Theorem: Under the null hypothesis that $x_t$ is an $ARIMA(p, 1, 0)$ process:

1. The asymptotic distribution of $t_{\hat{\phi}_{ols}}$ in regression (6) is the same as in the i.i.d. case.
2. $(\hat{\zeta}_1, \cdots, \hat{\zeta}_p)'$ are asymptotically normally distributed for any value of $d_1 \in [0, 1)$ used in the regression (6).

Theorem: Under the null hypothesis that $x_t$ is generated by

$$\Delta x_t = u_t, \qquad \alpha(B) u_t = a_t,$$

the asymptotic distribution of the t-ratio $t_{\hat{\phi}_{ols}}(\hat{d}_1)$ associated to coefficient $\phi$ in regression (6), where $\hat{d}_1$ is an estimator of $d_1$, as follows:

$$t_{\hat{\phi}_{ols}} \Rightarrow N(0, 1).$$

# Today's Topics

1. Long Memory Models
2. Seasonal Models

# Today's Topics

1. Long Memory Models

2. Seasonal Models
   Seasonality
   Seasonal Differencing
   Multiplicative Seasonal Models

# Seasonal Models

Some financial time series such as quarterly earnings per share of a company exhibits certain cyclical or periodic behavior. Such a time series is called a *seasonal time series*.

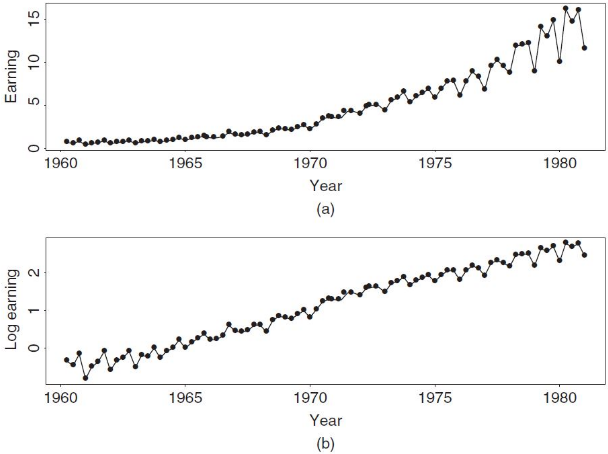


**Figure 2.13** Time plots of quarterly earnings per share of Johnson & Johnson from 1960 to 1980: (a) observed earnings and (b) log earnings.

In particular, the earnings grew exponentially during the sample period and had a strong seasonality. Furthermore, the variability of earnings increased over time. The cyclical pattern repeats itself every year so that the periodicity of the series is 4. If monthly data are considered (e.g., monthly sales of Wal-Mart stores), then the periodicity is 12. Seasonal time series models are also useful in pricing weather-related derivatives and energy futures because most environmental time series exhibit strong seasonal behavior.

# Applications

In some applications, seasonality is of secondary importance and is removed from the data, resulting in a seasonally adjusted time series that is then used to make inference. The procedure to remove seasonality from a time series is referred to as seasonal adjustment.

In other applications such as forecasting, seasonality is as important as other characteristics of the data and must be handled accordingly.

# Today's Topics

# Seasonal Differencing

Figure 2.13(b) shows the time plot of log earnings per share of Johnson & Johnson.

We took the log transformation for two reasons. First, it is used to handle the exponential growth of the series. Second, the transformation is used to stablize the variability of the series.

Log transformation is commonly used in analysis of financial and economic time series. In this particular instance, all earnings are positive so that no adjustment is needed before taking the transformation. In some cases, one may need to add a positive constant to every data point before taking the transformation.

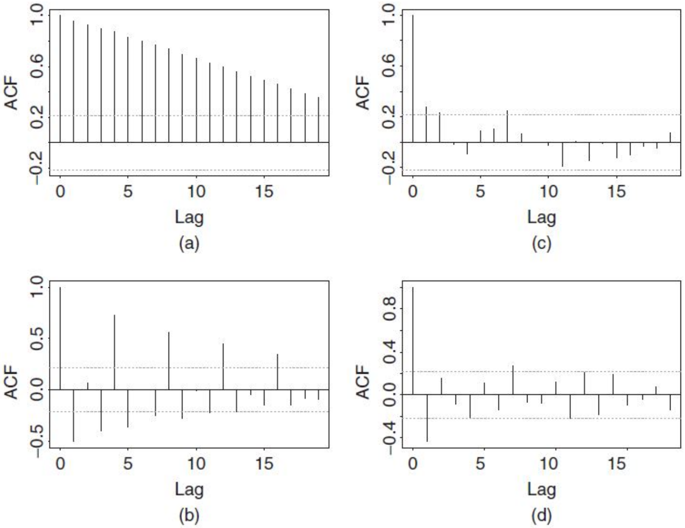


**Figure 2.14** Sample ACF of log series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980. (a) log earnings, (b) first differenced series, (c) seasonally differenced series, and (d) series with regular and seasonal differencing.

Box, Jenkins, and Reinsel (1994, Chapter 9):

$\Delta_4(\Delta x_t) = (1 - B^4)\Delta x_t = \Delta x_t - \Delta x_{t-4} = x_t - x_{t-1} - x_{t-4} + x_{t-5}$.

The operation $\Delta_4 = (1 - B^4)$ is called a *seasonal differencing*. In general, for a seasonal time series $y_t$ with periodicity $s$, seasonal differencing means

$$\Delta_s y_t = y_t - y_{t-s} = (1 - B^s) y_t.$$

The conventional difference $\Delta y_t = y_t - y_{t-1} = (1 - B) y_t$ is referred to as the *regular differencing*.

# Today's Topics

1. Long Memory Models

2. Seasonal Models
   - Seasonality
   - Seasonal Differencing
   - Multiplicative Seasonal Models

# Multiplicative Seasonal Models

A special seasonal time series model:

$$(1 - B^s)(1 - B)x_t = (1 - \theta B)(1 - \Theta B^s)a_t, \qquad (7)$$

where $s$ is the periodicity of the series, $a_t$ is a white noise series, $|\theta| < 1$, and $|\Theta| < 1$. This model is referred to as the *airline model* (Box,Jenkins, and Reinsel, 1994, Chapter 9).

The AR part of the model simply consists of the regular and seasonal differences, whereas the MA part involves two parameters.

Focusing on the MA part,

$$w_t = (1-\theta B)(1-\Theta B^s)a_t = a_t - \theta a_{t-1} - \Theta a_{t-s} + \theta\Theta a_{t-s-1},$$

where $w_t = (1-B^s)(1-B)x_t$ and $s > 1$. It is easy to obtain that $E(w_t) = 0$ and

$$\begin{aligned}
Var(w_t) &= (1+\theta^2)(1+\Theta^2)\sigma_a^2, \\
Cov(w_t, w_{t-1}) &= -\theta(1+\Theta^2)\sigma_a^2, \\
Cov(w_t, w_{t-s+1}) &= \theta\Theta\sigma_a^2, \\
Cov(w_t, w_{t-s}) &= -\theta(1+\Theta^2)\sigma_a^2, \\
Cov(w_t, w_{t-s-1}) &= \theta\Theta\sigma_a^2, \\
Cov(w_t, w_{t-l}) &= 0, \quad for \quad l \neq 0, 1, s-1, s, s+1.
\end{aligned}$$

Consequently, the ACF of the $w_t$ series is given by

$$\rho_1 = \frac{-\theta}{(1+\theta^2)}, \quad \rho_s = \frac{-\Theta}{(1+\Theta^2)}, \quad \rho_{s-1} = \rho_{s+1} = \rho_1\rho_s = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}$$

and $\rho_l = 0$ for $l > 0$ and $l \neq 1, s-1, s, s+1$.

It is interesting to compare the prior ACF with those of the $MA(1)$ model $y_t = (1 - B)a_t$ and the $MA(s)$ model $z_t = (1 - \Theta B^s)a_t$. The ACF of $y_t$ and $z_t$ series are

$$\begin{aligned} \rho_1(y) &= \frac{-\theta}{1+\theta^2} \qquad \textit{and} \qquad \rho_l(y) = 0,\ l > 1, \\ \rho_s(z) &= \frac{-\Theta}{1+\Theta^2} \qquad \textit{and} \qquad \rho_l(z) = 0,\ l > 0, \neq s. \end{aligned}$$

We see that (i) $\rho_1 = \rho_1(y)$, (ii) $\rho_s = \rho_s(z)$, and (iii) $\rho_{s-1} = \rho_{s+1} = \rho_1(y) \times \rho_s(z)$. Therefore, the ACF of $w_t$ at lags $(s-1)$ and $(s+1)$ can be regarded as the *interaction* between lag-1 and lag-$s$ serial dependence, and the model of $w_t$ is called a *multiplicative seasonal MA model*. In practice, a multiplicative seasonal model says that the dynamics of the regular and seasonal components of the series are approximately orthogonal.

The model

$$w_t = (1 - \theta B - \Theta B^s)a_t,$$

where $|\theta| < 1$ and $|\Theta| < 1$, is a nonmultiplicative seasonal MA model.

A multiplicative model is more parsimonious than the corresponding nonmultiplicative model because both models use the same number of parameters, but the multiplicative model has more nonzero ACFs.

# Example 1

we apply the airline model to the log series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980. Based on the exact-likelihood method, the fitted model is

$$(1-B)(1-B^4)x_t = (1-0.678B)(1-0.314B^4)a_t, \qquad \hat{\sigma}_a = 0.089,$$

where standard errors of the two MA parameters are 0.080 and 0.101, respectively. The Ljung-Box statistics of the residuals show $Q(12) = 10.0$ with a $p$ value of 0.44.

# Forecasting

Re-estimate the model using the first 76 observations and reserve the last 8 data points for forecasting evaluation.

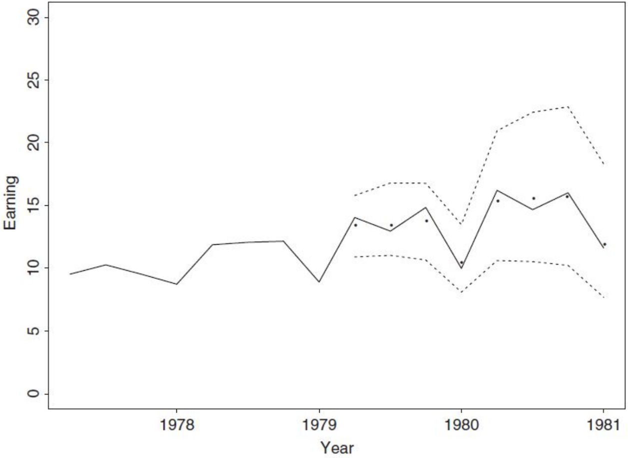


**Figure 2.15** Out-of-sample point and interval forecasts for quarterly earnings of Johnson & Johnson. Forecast origin is fourth quarter of 1978. In plot, solid line shows actual observations, dots represent point forecasts, and dashed lines show 95% interval forecasts.

# Deterministic v.s. Stochastic Seasonality

- When the seasonal pattern of a time series is stable over time (e.g., close to a deterministic function), dummy variables may be used to handle the seasonality.
- However, deterministic seasonality is a special case of the multiplicative seasonal model discussed before. Specifically, if $\Theta = 1$, then model (7) contains a deterministic seasonal component. Consequently, the same forecasts are obtained by using either dummy variables or a multiplicative seasonal model when the seasonal pattern is deterministic.
- Yet use of dummy variables can lead to inferior forecasts if the seasonal pattern is not deterministic. In practice, we recommend that the exact-likelihood method should be used to estimate a multiplicative seasonal model, especially when the sample size is small or when there is the possibility of having a deterministic seasonal component.

# Example 2

To demonstrate deterministic seasonal behavior, consider the monthly simple returns of the CRSP Decile 1 Index from January 1970 to December 2008 for 468 observations.

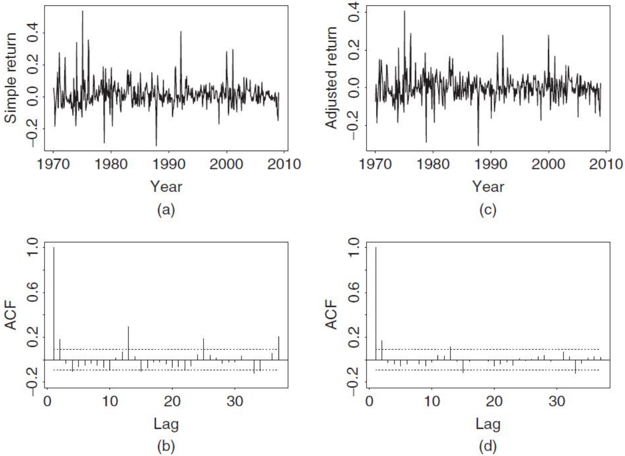


**Figure 2.16** Monthly simple returns of CRSP Decile 1 index from January 1970 to December 2008: (a) time plot of the simple returns, (b) sample ACF of simple returns, (c) time plot of simple returns after adjusting for January effect, and (d) sample ACF of adjusted simple returns.

If seasonal *ARMA* models are entertained, a model in the form

$$(1-\phi_1 B)(1-\phi_{12}B^{12})R_t = (1-\theta_{12}B^{12})a_t,$$

is identified, where $R_t$ denotes the monthly simple return. Using the conditionallikelihood method, the fitted model is

$$(1-0.18B)(1-0.87B^{12})R_t = (1-0.74B^{12})a_t, \qquad \tilde{\sigma}_a = 0.069.$$

If the exact-likelihood method is used, we have

$$(1-0.188B)(1-0.951B^{12})R_t = (1-0.997B^{12})a_t, \qquad \tilde{\sigma}_a = 0.063.$$

The estimation result suggests that the seasonal behavior might be deterministic. Thus, we define the dummy variable for January, that is,

$$Jan_t = \begin{cases} 1, & \text{if t is January;} \\ 0, & \text{otherwise.} \end{cases}$$

and employ the simple linear regression

$$R_t = \beta_0 + \beta_1 Jan_t + e_t.$$

The fitted model is $R_t = 0.0029 + 0.1253 Jan_t + e_t$, where the standard errors of the estimates are 0.0033 and 0.0115, respectively.

Consequently, the seasonal behavior in the monthly simple return of Decile 1 is mainly due to the *January effect* .