Financial Time Series

Topic 5: Determination of the order of integration of ARIMA models

Hung Chen
Department of Mathematics
National Taiwan University
4/12/2002
OUTLINE

1. Distinguishing between different values of \( d \)
2. Motivated Example
3. ACF of \( AR(p) \) with a Unit Root
4. Detection of Over-differencing
5. Testing for a Unit Root
   – The Dickey-Fuller Test
   – Non-parametric Test
6. Trend Stationarity vs Difference Stationarity
7. Trend Changes
   – Segmented trends
   – Structural breaks
   – Smooth transitions
ARIMA Models

Should we try a model other than ARMA?

General Wisdom:

- Consider a set of observation \( \{x_t, t = 1, 2, \ldots, n\} \).
- Suppose the data satisfies the following two characteristics:
  - It exhibits no apparent deviations from stationarity.
  - It has a rapidly decreasing autocorrelation function.

Then seek a suitable ARMA process to represent the mean-corrected data.

- Otherwise, first look for a transformation of the data which generates a new series with the above properties.

- A common transformation is **differencing**. It leads to the class of ARIMA processes.
  - The nonstationarity is mainly caused by the fact that there is no fixed level for price series.
– Such a nonstationary series is called unit-root time series.
   The best known example of unit-root time series is the random walk model.

• Question: How do we estimate the parameters of ARMA processes?
  – AR processes: The Yule-Walker Equation
  – MA processes: Use \( \rho_k \) and \( \text{Var}(X_t) \).
    We cannot get all consistent estimates of \( \rho_k \).
    Consider \( X_t = a_t + \theta a_{t-1} \).
    Then
    \[
    \begin{align*}
    \text{Var}(X_t) &= \sigma_a^2 + \theta^2 \sigma_a^2 \\
    \text{Cov}(X_t, X_{t-1}) &= \theta^2 \sigma_a^2.
    \end{align*}
    \]
  – \( ARM A(p, q) \) processes: Express it as an MA process and use the first \( p + q \rho_k \).
Motivated Example:

• Contrast between $I(0)$ and $I(1)$.

• $x_t \sim I(0)$ and assume that it has a zero mean.
  
  – The innovation $a_t$ has only a temporary effect on the value of $X_t$.
  
  – The variance of $X_t$ is finite and does not depend on $t$.
  
  – The expected length of times between crossings of $x = 0$ is finite.
  
  – The autocorrelation, $\rho_k$, decrease steadily in magnitude for large enough $k$, so that their sum is finite.

• $x_t \sim I(1)$ with $x_0 = 0$.
  
  – The innovation $a_t$ has a permanent effect on the value of $x_t$ because
    
    $$x_t = x_0 + \sum_{i=0}^{t} a_{t-i}.$$ 
  
  – The variance of $X_t$ goes to infinity as $t$ goes to infinity.
    
    $$\text{Var}(X_t) = \text{Var} \left( \sum_{i=0}^{t} a_{t-i} \right).$$
– The expected time between crossings of $x = 0$ is infinite.
– The autocorrelation, $\rho_k \to 1$ for all $k$ as $t$ goes to infinity.

- A time series is non-stationary is often self-evident from a plot of the series.
- Examination of the SACFs might be helpful to determine the actual form of non-stationarity.
ACF of $AR(p)$

- A stationary $AR(p)$ process requires that all roots with $|g_i| < 1$.

\[ \phi(B)X_t = a_t \]

\[ \phi(B) = (1 - g_1B)(1 - g_2B)\cdots(1 - g_pB). \]

- ACF:

\[ \rho_k = A_1g_1^k + A_2g_2^k + \cdots + A_pg_p^k. \]

- Random walk: $x_t = x_{t-1} + a_t$

- Random walk with drift: $x_t = x_{t-1} + \theta_0 + a_t$
  - $\theta_0$: the time-trend of the log price $x_t$.
  - It is often referred to as the drift of the model.
  - If we graph $x_t$ against time index $t$, we have a time-trend with slope $\theta_0$.

- Integrated processes: $\Delta x_t = \theta_0 + a_t$

- Suppose that one of $g_1, \ldots, g_p$ approaches 1.
  - $g_1 = 1 - \delta$, $\delta$: a small number
  - $\rho_k \approx A_1g_1^k$ since all other terms will go to zero more rapidly.
– Note that

\[ A_1g_1^k = A_1(1 - \delta)^k \cong A_1(1 - \delta k). \]

Failure of the SACF to die down quickly is an indication of non-stationarity.

• Possible strategy:

  – Suppose the original series \( x_t \) is found to be non-stationary, the first difference \( \Delta x_t \) is then analysed.
  – If \( \Delta x_t \) is still non-stationary, the next difference \( \Delta^2 x_t \) is then analysed.
  – Repeat this procedure until a stationary difference is found.
Detection of Over-differencing:

- Consider the stationary $MA(1)$ process $x_t = (1 - \theta B)a_t$.
- First difference:
  \[
  \triangle x_t = (1 - B)(1 - \theta B)a_t = (1 - \theta_1 B - \theta_2 B^2)a_t,
  \]
  where $\theta_1 + \theta_2 = (1 + \theta) - \theta = 1$.
- Non-invertible: $AR(\infty)$ representation does not exist. Estimation will be difficult.
- Variance:
  \[
  V(X_t) = (1 + \theta^2)\sigma^2 \\
  V(\triangle X_t) = 2(1 + \theta + \theta^2)\sigma^2.
  \]
- The variance of the overdifferenced process will be larger than that of the original process.
- The sample variance will decrease until a stationary sequence has been found, but will tend to increase on overdifferencing.
Testing for a Unit Root

- Consider the zero mean $AR(1)$ process with normal innovations
  \[
  x_t = \phi x_{t-1} + a_t, \quad t = 1, 2, \ldots, T \tag{1}
  \]
  where $a_t \sim NID(0, \sigma^2)$ and $x_0 = 0$.

- Suppose the process started at time $t = 0$ and $\phi > 1$. By (1),
  \[
  x_t = x_0 \phi^t + \sum_{i=1}^{t} \phi^i a_{t-i}.
  \]
  \[
  V(X_t) = \sigma^2 \frac{\phi^{2(t+1)} - 1}{\phi^2 - 1}
  \]
  \[
  E(X_t) = x_0 \phi^t \frac{\phi^{2(t+1)} - 1}{\phi^2 - 1}
  \]

- The OLS estimate of $\phi$ is given by
  \[
  \hat{\phi}_T = \frac{\sum_{t=1}^{T} x_{t-1} x_t}{\sum_{t=1}^{T} x_{t-1}^2}
  \]
  and
  \[
  \hat{\phi}_T - \phi = \frac{\sum_{t=1}^{T} x_{t-1} a_t}{\sum_{t=1}^{T} x_{t-1}^2}.
  \]

- When $|\phi| < 1$,
  \[
  \sqrt{T}(\hat{\phi}_T - \phi) \overset{d}{\sim} N \left(0, \frac{\sigma^2}{EX_t^2} \right).
  \]
• Note that

\[ E(X^2_{t-1}) = E\left( \sum_{i=0}^{\infty} \phi^i a_{t-i} \right)^2 \]
\[ = \sigma^2/(1 - \phi^2). \]

Hence, \( \sqrt{T}(\hat{\phi}_T - \phi) \sim N(0, 1 - \phi^2) \).

• When \( \phi = 1 \), the above result breaks down.

What is the right distribution of \( \hat{\phi}_T - \phi \) under suitable normalization when \( \phi = 1 \)?

Write

\[ T(\hat{\phi}_T - \phi) = \frac{T^{-1} \sum_{t=1}^{T} x_{t-1} a_t}{T^{-2} \sum_{t=1}^{T} x_{t-1}^2}. \] (2)
What is $T^{-1} \sum_{t=1}^{T} x_{t-1} a_t$?

- When $\phi = 1$, $x_t = \sum_{s=1}^{t} a_s$ and hence $x_t \sim N(0, \sigma^2 t)$.
- Note that
  \[
  x_{t-1} a_t = (x_t^2 - x_{t-1}^2 - a_t^2)/2
  \]
  and
  \[
  \sum_{t=1}^{T} x_{t-1} a_t = \frac{x_T^2 - x_0^2}{2} - \frac{1}{2} \sum_{t=1}^{T} a_t^2.
  \]
- Recall that $x_0 = 0$ and hence
  \[
  \frac{1}{\sigma^2 T} \sum_{t=1}^{T} x_{t-1} a_t = \frac{1}{2} \left( \frac{x_T}{\sigma \sqrt{T}} \right)^2 - \frac{1}{2} \frac{1}{\sigma^2 T} \sum_{t=1}^{T} a_t^2.
  \]
- $x_T/(\sigma \sqrt{T})$ is $N(0, 1)$.
- $T^{-1} \sum_{t=1}^{T} a_t^2$ converges in probability to $\sigma^2$.
- Thus
  \[
  T^{-1} \sum_{t=1}^{T} x_{t-1} a_t \overset{a}{\sim} (1/2)\sigma^2(X - 1)
  \]
  where $X \sim \chi_1^2$. 

11
What is $T^{-2} \sum_{t=1}^{T} x_{t-1}^2$?

- Why do we consider $T^{-2}$ normalization?

\[
E[\sum_{t=1}^{T} X_{t-1}^2] = \sigma^2 \sum_{t=1}^{T} (t - 1) = \sigma^2(T - 1)T/2
\]

and

\[
E[T^{-2} \sum_{t=1}^{T} X_{t-1}^2] \to \sigma^2/2.
\]

- Denote $[rT]$ as the integer part of $rT$, $0 \in [0, 1]$, and define the random step function $R_T(r)$ as follows.

\[
R_T(r) = x_{[rT]}(r)/\sigma \sqrt{T}.
\]

- Properties of $R_T(r)$:

  - $[0, 1]$ is divided into $T + 1$ parts at $r = 0, T^{-1}, \ldots, 1$.
  
  - $R_T(r)$ is constant at values of $r$ but with jumps at successive integers.
  
  - As $T \to \infty$, $R_T(r)$ weakly converges to standard Brownian motion (or the Wiener process), $W(r)$, denoted

\[
R_T(r) \Rightarrow W(r) \sim N(0, r).
\]
• **Standard Brownian Motion:**
  It starts at level zero and satisfies the conditions
  
  $- W(0) = 0,$
  
  $- W(r_2) - W(r_1), W(r_3) - W(r_2), \ldots, W(r_n) - W(r_{n-1})$ are independent for every $n \in \{3, 4, \ldots\}$ and every $0 \leq r_1 < \ldots < r_n,$
  
  $- W(r) - W(s) \sim N(0, r - s)$ for $r \geq s.$

• **Facts:**

  $$W^2(1) - 1 = 2 \int_0^1 W(r) dW(r)$$
  
  $$W(1) \sim N(0, 1)$$
  
  $$\sigma \cdot W(r) \sim N(0, \sigma^2 r)$$
  
  $$W^2(r)/r \sim \chi^2_1$$
  
  $$f(R_T(r)) \Rightarrow f(W(r))$$

  if $f(\cdot)$ is a continuous functional on $[0, 1].$

• **Observe that**

  $$T^{-2} \sum_{t=1}^{T} x_{t-1}^2 = \sigma^2 T^{-1} \sum_{t=1}^{T} \left( \frac{x_{t-1}}{\sigma \sqrt{T}} \right)^2$$
  
  $$= \sigma^2 \sum_{t=1}^{T} T^{-1} \left( R_T((t - 1)/T) \right)^2$$
  
  $$= \sigma^2 \sum_{t=1}^{T} \int_{(i-1)/T}^{i/T} R^2_T(r) dr$$
\[ \rightarrow \sigma^2 \int_0^1 W^2(r)dr. \]

- Note that
\[ T^{-1} \sum_{t=1}^T X_{t-1}a_t \rightarrow \frac{\sigma^2}{2}(W^2(1) - 1). \]

- We conclude that
\[ T(\hat{\phi}_T - 1) \Rightarrow \frac{[W^2(1) - 1]/2}{\int_0^1 W^2(r)dr}. \quad (3) \]

Why does \( W^2(1) - 1 = 2 \int_0^1 W(r)dW(r) \) hold?

- The sample path of \( W(r) \) is almost uniformly continuous.

- Almost every Brownian path is nowhere differentiable.

- Define \( \int_0^1 f(r)dW(r) \) as
\[ \lim_{\epsilon \to 0} \int_0^1 f(r) \frac{W(r + \epsilon) - W(r)}{\epsilon} dr. \]
Here \( f \) is continuously differentiable.

Note that
\[ \int_0^1 f(r) \frac{W(r + \epsilon) - W(r)}{\epsilon} dr \]
\[ = \int_0^1 f(r) \frac{d}{dr} \left( \frac{1}{\epsilon} \int_r^{r+\epsilon} W(s)ds \right) dr. \]
Apply the integration by parts, we have
\[ \int_0^1 f(r) \frac{W(r + \epsilon) - W(r)}{\epsilon} dr \]
\[ \rightarrow \left[ f(r) \frac{1}{\epsilon} \int_r^{r+\epsilon} W(s) ds \right]_{0}^{1} \]
\[ - \int_0^1 \left( \frac{1}{\epsilon} \int_r^{r+\epsilon} W(s) ds \right) df(r) \]
\[ = f(1)W(1) - f(0)W(0) - \int_0^1 W(r)df(r). \]

- \( W(r) \) is not differentiable. Suppose we just plug \( W \) to the above formular, we have
\[ \int_0^1 W(r)dW(r) = \frac{1}{2}W^2(1). \]

How do we handle it?
Alternative test: the $t$-statistic

Test $\phi = 1$.

- The $t$-statistic

$$t_\phi = \tau = \frac{\hat{\phi}_T - 1}{\hat{\sigma}_{\hat{\phi}_T}}$$  \hspace{1cm} (4)

where

$$\hat{\sigma}_{\hat{\phi}_T} = \left( \frac{s_T^2}{\sum_{t=1}^{T} x_{t-1}^2} \right)^{1/2}$$

and

$$s_T^2 = \sum_{t=1}^{T} \frac{(x_t - \hat{\phi}_T x_{t-1})}{(T - 1)}.$$

- By (2) and (4), we have

$$\tau = \frac{T^{-1} \sum_{t=1}^{T} x_{t-1} a_t}{s_T / \sigma (\sum_{t=1}^{T} x_{t-1}^2)^{1/2}}.$$

- $s_T^2$ is a consistent estimator of $\sigma^2$.

- By the above argument, we have

$$\tau \Rightarrow \frac{\sigma^2 (W^2(1) - 1)/2}{\sigma (\sigma^2 \int_0^1 W^2(r) dr)^{1/2}}$$

$$= \frac{\int_0^1 W(r) dW(r)}{(\int_0^1 W^2(r) dr)^{1/2}}.$$ \hspace{1cm} (5)
• Dickey-Fuller test

Use Monte Carlo simulation to find the limiting distribution of (3)

• Recall that \( x_t = \sum_{s=1}^{t} a_s \)

• Simulate \( a_t \) by drawing \( t \) pseudo-random \( N(0,1) \) variates.

• Calculate

\[
T \frac{\sum_{t=1}^{T} \left( \sum_{s=0}^{t-1} a_s \right) a_t}{\sum_{t=1}^{T} \left( \sum_{s=0}^{t-1} a_s \right)^2}.
\]

• Repeat this calculation \( n \) times and compile the results into an empirical probability distribution.

• Refer to page 71 on discussions related to this topic.
Extensions to the Dickey-Fuller test

Extension 1:
Consider

\[ x_t = \theta_0 + \phi x_{t-1} + a_t, \quad t = 1, 2, \ldots, T, \] (6)

in which the mean may not be zero.

- Note that the unit root null is parametrized as \( \theta_0 = 0 \) and \( \phi = 1 \) in (6).

The tests have to be modified as follows.

\[
T(\hat{\phi}_T - 1) \Rightarrow \frac{[W^2(1) - 1]/2 - W(1) \cdot \int_0^1 W(r)dr}{\int_0^1 W^2(r)dr - (\int_0^1 W(r)dW(r))^2},
\]

\[
\tau_{\mu} \Rightarrow \frac{[W^2(1) - 1]/2 - W(1) \cdot \int_0^1 W(r)dr}{\left\{ \int_0^1 W^2(r)dr (\int_0^1 W(r)dW(r))^2 \right\}^{1/2}}.
\]

- Wald test:

  - restricted residual sum of squares:

    \[
    \sum_{t=1}^T (\Delta x_t)^2
    \]

  - unrestricted residual sum of squares:

    \[
    \sum_{t=1}^T \hat{\alpha}_t^2 = \sum_{t=1}^T (x_t - \hat{\theta}_0 - \hat{\phi}_T x_{t-1})^2
    \]
- Test statistic:
\[ \Phi = \frac{(\Sigma_{t=1}^{T}(\Delta x_t)^2 - \Sigma_{i=1}^{T} \hat{a}_i^2)}{\Sigma_{i=1}^{T} \hat{a}_i^2/(T - 2)}. \]

- The limiting distribution is tabulated in Dickey and Fuller (1981).

- The above distribution results are still valid as long as \( T \) is large and the innovations have finite variances.

Extension 2:

Consider the \( AR(p) \) process
\[ (1 - - \cdots - \phi_p B^p)x_t = \theta_0 + a_t \]
or
\[ x_t = \theta_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + a_t. \] (7)

Define
\[ \phi = \sum_{i=1}^{p} \phi_i \]
\[ \delta_i = - \sum_{j=i+1}^{p-1} \phi_j, \quad i = 1, 2, \ldots, p - 1. \]

Rewrite (7) as
\[ x_t = \theta_0 + \phi x_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta x_{t-i} + a_t. \] (8)

The A(ugmented)DF test:
• The null of one unit root is $\phi = \sum_{i=1}^{p} \phi_i = 1$.

• The test

$$\tau_\mu = \frac{\hat{\phi}_T - 1}{se(\hat{\phi}_T)}$$

where $se(\hat{\phi}_T)$ is the OLS standard error attached to the estimate $\hat{\phi}_T$.

• The above test has the same limiting distribution as

• $T(\hat{\phi}_T - 1)$ and the Wald $\Phi$ test have identical distributions to those obtained in the $AR(1)$ case.

• Refer to page 74 for discussions related to $ARMA(p, q)$ and $p$ and $q$ are unknown.
Non-parametric Tests:
remove white noise assumption

Consider the model

\[ x_t = \theta_0 + \phi x_{t-1} + a_t, \quad t = 1, \ldots, T. \quad (9) \]

Assumptions on \( \{a_t\}_1^\infty \):

- \( E(a_t) = 0 \) for all \( t \);
- \( \sup_t E(|a_t|^\beta) < \infty \) for some \( \beta > 2 \);
- \( \sigma^2_s = \lim_{T \to \infty} E(T^{-1}S_T^2) \) exists and is positive, where \( S_T = \sum_{t=1}^{T} a_t \);
- \( a_t \) is strong mixing, with mixing numbers \( \alpha_m \) that satisfy \( \sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty \).

Remarks:

- The above assumptions will be referred later as Assumption I.

- Allow heterogeneity.

- The third one is to ensure non-degenerate limiting distributions.

- The mixing numbers \( \alpha_m \) measure the strength and extent of temporal dependence within the sequence \( a_t \).
• The fourth one ensures that $a_t$ is weakly dependent. Dependence declines as the length of memory ($m$) increases.

• If $a_t$ is stationary, then

$$\sigma_S^2 = E(a_t^2) + 2 \sum_{j=2}^{\infty} E(a_1 a_j).$$

• If $a_t$ is the MA(1) process ($a_t = \epsilon_t - \theta \epsilon_{t-1}$), then

$$\sigma_S^2 = \sigma_\epsilon^2 (1 + \theta^2) - 2 \sigma_\epsilon^2 \theta = \sigma_\epsilon^2 (1 - \theta)^2.$$  

• If $a_t$ is white noise, $\sigma_S^2 = \sigma_\epsilon^2$.

• Define

$$\sigma^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E(a_t^2).$$

How do we handle $\sigma^2 \neq \sigma_S^2$?

Consider the following asymptotically valid test.

$$Z(\phi) = T(\hat{\phi} - 1) - \frac{\hat{\sigma}_{S\ell}^2 - \hat{\sigma}^2}{2}$$

$$\cdot \left[ T^{-2} \sum_{t=2}^{T} (x_{t-1} - \bar{x}_{-1})^2 \right]^{-1}.$$  

Here
\[ \bar{x}_{-1} = (T - 1)^{-1} \sum_{t=1}^{T-1} x_t \]

\[ \hat{\sigma}^2_{S\ell} \text{ is} \]
\[ T^{-1} \sum_{t=1}^{T} \hat{\alpha}_t^2 + 2T^{-1} \sum_{j=1}^{\ell} \sum_{t=j+1}^{T} \hat{\alpha}_t \hat{\alpha}_{t-j}. \]

• The lag truncation parameter \( \ell \) can be set to be \([T^{0.25}]\).

• Let \( \hat{\alpha}_t \) be the residual from estimating (9). Then
\[ \hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{\alpha}_t^2. \]

Another asymptotically valid test:
\[ Z(\tau_\mu) = \tau_\mu \left( \hat{\sigma}^2 / \hat{\sigma}^2_{S\ell} \right) - \frac{\hat{\sigma}^2_{S\ell} - \hat{\sigma}^2}{2} \]
\[ \cdot T \left[ \hat{\sigma}^2_{S\ell} \sum_{t=2}^{T} (x_{t-1} - \bar{x}_{-1})^2 \right]^{-1/2}. \]

Under the unit root null, the above two statistics have the same limiting distributions as \( T(\hat{\phi}_T - 1) \) and \( \tau_\mu \), respectively.
More Than One Unit Root

- Sometimes, we need differencing twice to induce stationarity.
- The Dickey-Fuller type tests are based on the assumption of at most one unit root.

Proposed procedure:

- Test $H_0$: two unit roots against $H_a$: one unit root, consider
  \[ \Delta^2 x_t = \beta_0 + \beta_2 \Delta x_{t-1} + a_t. \]

- Compare the t-ratio on $\beta_2$ from the above regression with the $\tau_\mu$ critical values.

- If the null is rejected, we then test $H_0$: one unit root against $H_a$: no unit root. Consider
  \[ \Delta^2 x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 \Delta x_{t-1} + a_t. \]

- Compare the t-ratio on $\beta_1$ from the above regression with the $\tau_\mu$ critical values.
Stochastic unit root processes (STUR)

- Random Coefficient $AR(1)$ process
  \[ x_t = \phi_t x_{t-1} + a_t, \quad \phi_t = 1 + \delta_t \]
  where $a_t$ and $\delta_t$ are independent zero mean white-noise processes with variances $\sigma_a^2$ and $\sigma_{\delta}^2$.

- Motivation for STUR:
  - $x_t$: the price of a financial asset.
  - Consider the expected return at time $t$
    \[ E(r_t) = \frac{E(x_t) - x_{t-1}}{x_{t-1}}. \]
    For simplicity, dividend payments are ignored.
  - $E(x_t) = (1 + E(r_t))x_{t-1}$.
  - Set $a_t = x_t - E(x_t)$ and $\delta_t = r_t$.
  - The price levels have a stochastic unit root.

- Alternative formulation considered in Granger and Swanson (1997):
  \[ \phi_t = \exp(\alpha_t) \]
where $\alpha_t$ is a zero mean stationary stochastic process.
Granger and Swanson’s Model

Recall $\phi_t = \exp(\alpha_t)$.

- $\phi_t = (x_t/x_{t-1})(1 - a_t/x_t)$
  
  Observe that

  $$\alpha_t = \Delta \log(x_t) + \log(1 - a_t/x_t)$$

  $$\approx \Delta \log(x_t) - a_t/x_t.$$

- $\log(x_t)$ has an exact unit root and $x_t$ has a stochastic unit root.

- The daily levels of the London Stock Exchange FTSE 350 index over the period 1 January 1986 to 28 November 1994 is fitted by the following $STUR(4)$ model

  $$\Delta x_t = \beta + \phi_1 \Delta x_{t-1} + \phi_4 \Delta x_{t-4}$$

  $$+ \delta_t[x_{t-1} - \beta(t - 1) - \phi_1 x_{t-2} + \phi_4 x_{t-5}]$$

  $$+ a_t$$

  $$\delta_t = \delta_{t-1} + \eta_t.$$
Trend stationarity versus difference stationarity

Efficient Market Hypothesis:

- When prices follow a random walk (unit root) the only relevant information in the series of present and past prices, for trader, is the most recent price.
- In the above case, the people involved in the market have already made perfect use of the information in past prices.
- A market will be called perfectly efficient if the prices fully reflect available information, so that prices adjust fully and instantaneously when new information becomes available.

Unit root testing strategy:

- Null hypothesis: The series is generated as a driftless random walk with, possibly, a serially correlated error.
- The null hypothesis is called difference stationary in Nelson and Plosser (1982).

\[ \Delta x_t = \epsilon_t, \]  

(11)
where $\epsilon_t = \theta(B)a_t$.

- This null hypothesis is appropriate for financial time series such as interest rates and exchange rates.

- The alternative is that $x_t$ is stationary in levels.

Another setting:

- Many financial time series contain a drift.

- The null hypothesis:

$$\Delta x_t = \theta + \epsilon_t.$$  \hspace{1cm} (12)

- The alternative hypothesis:

$$x_t = \beta_0 + \beta_1 t + \epsilon_t.$$  \hspace{1cm} (13)

$x_t$ is generated by a linear trend buried in stationary noise.

It is **trend stationary** (TS).

Consider an AR type of model ($t$: additional regressor)

$$x_t = \beta_0 + \beta_1 t + \phi x_{t-1} + \sum_{i=1}^{k} \delta_i \Delta x_{t-i} + a_t$$  \hspace{1cm} (14)
and the statistic
\[
\tau = \frac{\hat{\phi}_T - 1}{se(\hat{\phi}_T)}.
\]

Its limiting distribution is
\[
\frac{[W^2(1) - 1]/2 - W(1) \int_0^1 W(r)dr + A}{\left\{ \int_0^1 W^2(r)dr - (\int_0^1 W(r)dr)^2 + B \right\}^{1/2}}
\]
where
\[
A = 12 \left[ \int_0^1 rW(r)dr - (1/2) \int_0^1 W(r)dr \right] \times \left[ \int_0^1 W(r)dr - W(1)/2 \right]
\]
and
\[
B = 12 \left[ \int_0^1 W(r)dr \int_0^1 rW(r)dr - (\int_0^1 rW(r)dr)^2 \right] - 3 \left( \int_0^1 W(r)dr \right)^2.
\]

Refer to page 81 for the non-parametric test statistic.

Problem with (14)

Question: If \( \beta_1 \neq 0 \), \( x_t \) will contain a quadratic trend.
• Consider $p = 1$, (14) can be written as

$$x_t = \beta_0 \sum_{j=1}^{t} \phi^{t-j} + \beta_1 \sum_{j=1}^{t} j\phi^{t-j} + \sum_{j=1}^{t} a_j \phi^{t-j}.$$  

• Under the null $\phi = 1$,

$$x_t = \beta_0 t + \beta_1 t(t + 1)/2 + S_t.$$  

• Quadratic trend is unlikely because a non-zero $\beta_1$ under the null would imply an ever-increasing (or decreasing) rate of change $\Delta x_t$. 

31
Trend Changes

We just consider whether the observed series \( \{x_t\}_{0}^{T} \) is a realization from a process characterized by the presence of a \textbf{unit root} and possibly a non-zero \textbf{drift}.

Perron’s (1989) Suggestion:

One-Time Change in the structure at time \( T_B \)

Idea: an exogenous change in the level of the series

How do we accommodate this change?

We first consider segmented trends.

Model A:

- Example: Consider S&P stock index which goes through the Great Crash of 1929. \( T_B = 1929 \).
  
  Refer to Figure 3.7 for further detail.

- \( x_t = \mu + x_{t-1} + bDTB_t + e_t \).

- \( DTB_t = 1 \) if \( t = T_B + 1 \) and 0 otherwise.

- \( e_t \) satisfies Assumption I.

- Model A characterizes the \textbf{crash} by a dummy variable which takes the value one at the
time of the break.
After the crash, it resumes to the normal.

● Possible alternative: Consider

\[ x_t = \mu_1 + \beta t + (\mu_2 - \mu_1)DU_t + e_t, \]

where \( DU_t = 1 \) if \( t > T_B \) and 0 otherwise.

● The above alternative means that a one-
time change in the intercept of the trend
function.
The magnitude change is \( \mu_2 - \mu_1 \).

Model B:

● Figure 3.7 suggests the possibility of both a
change in level and, thereafter, an increased
trend rate of growth of the series.

● \( x_t = \mu_1 + x_{t-1} + (\mu_2 - \mu_1)DU_t + e_t. \)

● Model B (changing growth model) assumes
that the drift parameter changes from \( \mu_1 \) to
\( \mu_2 \) at time \( T_B \).

● Possible alternative: Consider

\[ x_t = \mu_1 + \beta_1 t + (\beta_2 - \beta_1)DT_t^* + e_t, \]

where \( DT_t^* = t - T_B \) if \( t > T_B \) and 0 oth-
otherwise.
• The above alternative means that a change in the slope of the trend function (of magnitude $\beta_2 - \beta_1$), without any sudden change in the level.

Model C:

• Figure 3.7 suggests that a sudden change in the level followed by a different growth.

• $x_t = \mu_1 + x_{t-1} + \zeta DT B_t + (\mu_2 - \mu_1) D U_t + e_t$.

• Model C assumes that a sudden change followed by the drift parameter change from $\mu_1$ to $\mu_2$ at time $T_B$.

• Possible alternative: Consider

$$x_t = \mu_1 + \beta_1 t + (\mu_2 - \mu_1) D U_t + (\beta_2 - \beta_1) D T^*_t + e_t.$$

• The above alternative allows both effects to take place simultaneously.

a sudden change in the level followed by a different growth path
Multiple Structure Break

Logistic smooth transition regression (LSTR): Allow the trend to change gradually and smoothly between two regimes.

- Three Models:
  Model A:
  \[ x_t = \mu_1 + \mu_2 S_t(\gamma, m) + e_t. \]
  Model B:
  \[ x_t = \mu_1 + \beta_1 t + \mu_2 S_t(\gamma, m) + e_t. \]
  Model C:
  \[ x_t = \mu_1 + \beta_1 t + \mu_2 S_t(\gamma, m) + \beta_2 t S_t(\gamma, m) + e_t. \]
- The logistic smooth transition function
  \[ S_t(\gamma, m) = (1 + \exp(-\gamma(t - mT)))^{-1}. \]
- \( m \): the timing of the transition midpoint,
  \( S_{mT}(\gamma, m) = 0.5. \)
- \( \gamma \): the speed of transition
  For \( \gamma > 0, \)
  \[ S_{-\infty}(\gamma, m) = 0, \quad S_\infty(\gamma, m) = 1. \]
As $\gamma \to \infty$, $S_t(\gamma, m)$ changes from 0 to 1 instantaneously at time $mT$.

- Model A: $x_t$ is stationary around a mean which changes from $\mu_1$ to $\mu_1 + \mu_2$.
- Model B: The intercept changes from $\mu_1$ to $\mu_1 + \mu_2$ but allows for a fixed slope.
- Model C: The intercept changes from $\mu_1$ to $\mu_1 + \mu_2$ and the slope also changes from $\beta_1$ to $\beta_1 + \beta_2$. 