Introduction to Time Series Analysis: Stochastic Process

Class 8

Wonmun Shin

(wonmun.shin@sejong.ac.kr)

Department of Economics, Sejong University

* This lecture note is written based on Professor Chang-Jin Kim's lecture note and Gujarati textbook (Chapter 21, 5th edition).

Stochastic Process

Introduction: Cross-Sectional Data vs. Time Series Data

- **Cross-sectional data**: data on one or more variables collected *at the same point in time*
  - Population and housing census by the Statistics Korea every 5 years
  - Consumer survey index by Bank of Korea every quarter
  - Opinion polls by the Gallup Korea
- **Time series data**: a set of observations on the values that a variable takes *at different times*
  - Daily data: stock prices, weather reports
  - Monthly data: unemployment rate, CPI (Consumer Price Index)
  - Annual data: GDP

Stochastic Process

- **Stochastic process**: *a collection of random variables ordered in time*

$$Y_1,\ Y_2,\ Y_3,\ Y_4,\ \cdots,\ Y_T$$

  - Keep in mind that each of Y's is a random variable!
  - For example, Y_t is GDP of the year t. (Y_1 = GDP of 1960, Y_2 = GDP of 1961, $\cdots$)

- The term "stochastic" comes from the Greek word "*stokhos*".
  - *Stokhos* means a bull's-eye (or a center of a target).
  - If you have ever thrown darts on a dart board with the aim of hitting the bull's-eye, how often did you hit the bull's-eye?
  - Out of a hundred darts you may be lucky to hit the bull's-eye only a few times. At other times the darts will be spread randomly around the bull's-eye.

- Going back to the example of GDP:
  - Y_t is GDP of the year t. (Y_1 = GDP of 1960, Y_2 = GDP of 1961, $\cdots$)
  - We know that the GDP of 2019 was 1,919 trillion won.
  - Theoretically, the GDP figure for 2019 could have been any number, depending on the economic and political circumstances. $\rightarrow$ The figure of 1,919 trillion won is a particular realization of all such possibilities.
  - In summary, we can say that GDP is a **stochastic process**, and the actual values we observe for the sample period are particular **realization** of the process.

Stochastic Process [cont'd]

- The distinction between the stochastic process and its realization is akin to the distinction between population and sample in cross-sectional data.
- **However** the critical difference between "realization of stochastic process" and "sample of population" is *whether we can draw it once again or not.*
  - We cannot observe another possibility of Y_t except the actual realization (unless we have a time machine).
  - That is, we do not have any information about Y_t other than the realization.
- We want to analyze with a stochastic process (Y_t for $t = 1, 2, \cdots, T$), but how if we observe only one realization for each random variable?
  - Suppose we want to forecast Y_{T+1} (out-of-sample) $\rightarrow$ We need to know $E(Y_t)$ and $Var(Y_t)$.
  - But the expectations and variances for each Y_t can be different!
  - We can't get anything from time series data unless we assume about Y_t.

Stationary Stochastic Process

Stationary Stochastic Process

- *"A **stochastic process** is said to be **stationary** if its **mean** and **variance** are **constant over time** and the value of the **covariance** between two time periods **depends only on the distance** or gap or lag between two time periods and not the actual time at which the covariance is computed."* - *Gujarati and Porter, Basic Econometrics, 5th edition, McGraw-Hill*

- **Stationarity**: The time series Y_t is said to be stationary if

$$E(Y_t) = \mu, \quad \forall t$$
$$Var(Y_t) = \sigma^2, \quad \forall t$$
$$Cov(Y_t, Y_{t+k}) = \gamma_k, \quad \forall t, k$$

  - The time series satisfying the above condition is known as a *weakly stationary* process.
  - It is also known as a *covariance stationary* process, or a *second-order stationary* process, or a *wide-sense stationary* process.

Stationary Stochastic Process

- If a time series is stationary, its mean, variance, and autocovariance remain the same no matter what point we measure them.
- That is, the first and second moments of a time series are **time invariant**!
  - Such a time series will tend to return to its mean (called *mean reversion*).
  - Fluctuations around the mean (measured by the variance) will have a broadly constant amplitude.
  - The speed of mean reversion depends on the autocovariances; it is quick if the autocovariances are small and slow when they are large.
- If a time series is not stationary in the sense we defined, it is called a **non-stationary time series**.
  - In other words, a non-stationary process will have a time-varying mean or a time-varying variance or both.

Stationary Stochastic Process [cont'd]

- *(optional)* A time series is *strictly stationary* if all the moments of its probability distribution are invariant over time.
  - However, if the stationary process is normal, the weakly stationary process is also strictly stationary.
  - Why? The normal stochastic process is fully specified by its first and second moments, the mean and the variance.
  - In that sense, again, the central limit theorem (CLT) is quite important!

Non-Stationarity

- Classical example of non-stationary process is a **random walk**.
  - It is often said that asset prices such as stock prices or exchange rates follow a random walk.
  - We distinguish two types of random walks: (i) random walk without drift, (ii) random walk with drift

- **Random walk *without drift***

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

  - e_t is a (Gaussian) **white noise** error term with mean 0 and variance σ^2: $e_t \sim iidN(0, \sigma^2)$
  - The value of Y at time t is equal to its value at time $t-1$ plus a random shock.

Random Walk without Drift [cont'd]

- We can write:

$$Y_1 = Y_0 + e_1$$
$$Y_2 = Y_1 + e_2 = Y_0 + e_1 + e_2$$
$$Y_3 = Y_2 + e_3 = Y_o + e_1 + e_2 + e_3$$

- In general, we have

$$Y_t = Y_0 + \Sigma e_t$$

- Therefore,

$$E\left(Y_t\right) = E\left(Y_0 + \Sigma e_t\right) = Y_0$$
$$Var\left(Y_t\right) = Var\left(Y_0 + \Sigma e_t\right) = t\sigma^2$$

  - The mean of Y is equal to its initial value which is constant.
  - But as t increases, its variance indefinitely increases. $\rightarrow$ A random walk without drift is non-stationary.

Random Walk without Drift [cont'd]

- An interesting feature of the random walk is the *persistence of random shocks*.
  - Y_t is initial Y_0 plus the sum of random shocks.
  - As a result, the impact of a particular shock does not die away.
  - This is why random walk is said to have an *infinite memory*.
- Interestingly, if we write the equation (1) as

$$Y_t - Y_{t-1} = e_t$$
$$\rightarrow \quad \triangle Y_t = e_t$$

  - $\triangle$ is the first difference operator.
  - While Y_t is nonstationary, $\triangle Y_t$ is stationary!
  - That is, the first differences of a random walk time series are stationary.

Random Walk with Drift

- Let's modify the equation (1) as follows:

$$Y_t = \delta + Y_{t-1} + e_t \tag{2}$$

  - δ is known as the **drift parameter**.

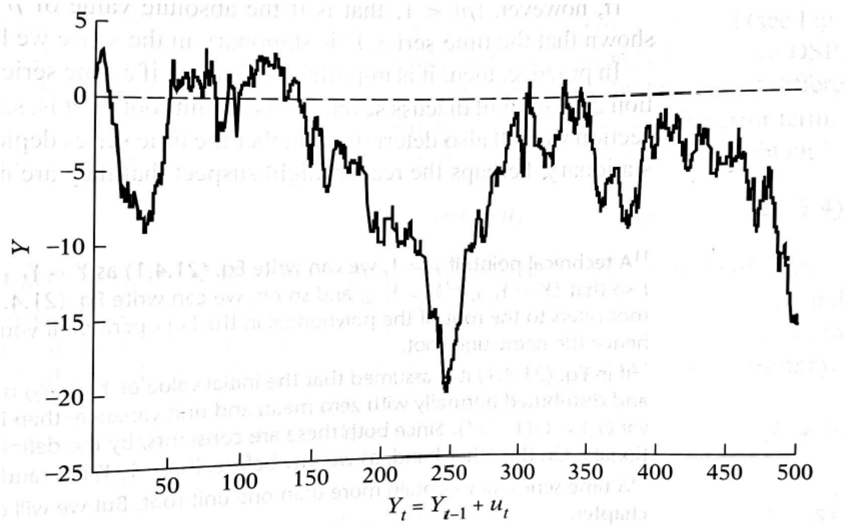
- We can show that

$$E\left(Y_t\right) = Y_0 + t\delta$$
$$Var\left(Y_t\right) = t\sigma^2$$

  - For a random walk with drift, the mean as well as the variance increase over time, which violates the conditions of stationarity.
  - In short, a **random walk**, with or without drift, is a **non-stationary stochastic process**.

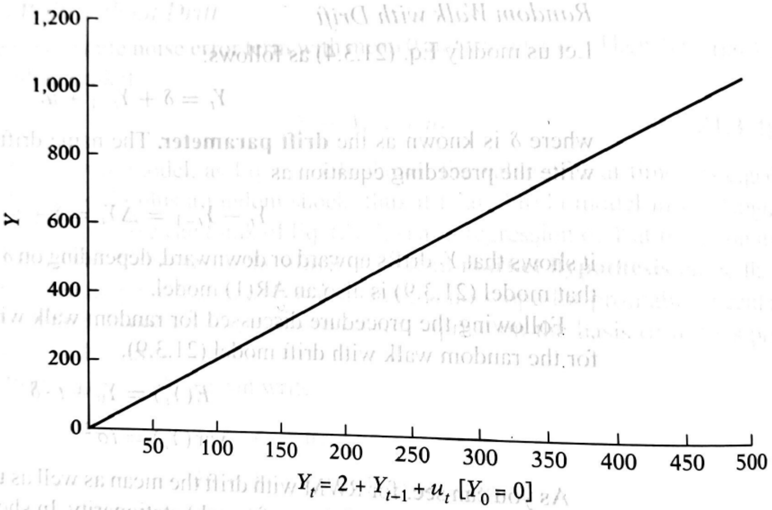
Random walk without drift (DGP: $Y_t = Y_0 + e_t,\ Y_0 = 0,\ e_t \sim N(0,1)$)

**source: Gujarati Figure 21.3*



Non-Stationary Stochastic Process: Random Walk [cont'd]

Random walk with drift (DGP: $Y_t = \delta + Y_0 + e_t$, $\delta = 2$, $Y_0 = 0$, $e_t \sim N(0,1)$)
**source: Gujarati Figure 21.4*



Unit Root Process

- The random walk model is an example of what is known as a **unit root process**.
- Let us write the random walk model (Equation 1) as:

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

  - This model resembles the first-order autoregressive model (that we discussed in the class of autocorrelation).
  - If $-1 < \rho < 1$, the process has zero mean, homoskedastic variance, and the covariance only depends on the distance of two time periods. $\rightarrow$ Stationarity!
  - If $\rho = 1$, the process becomes a random walk. $\rightarrow$ Non-stationarity!
- Therefore, if $\rho = 1$, we face the **unit root problem**.
  - The name *"unit root"* is due to the fact that $\rho = 1$.
  - Thus the terms *"non-stationarity"*, *"random walk"*, *"unit root"* (and *"stochastic trend"* which we will learn soon) can be treated synonymously.

Unit Root Test

- In practice, then, it is important to find out if a time series possesses a unit root.
- In this class, we will introduce the idea of **unit root test** (which is widely popular test of stationarity).
  - We will revisit the unit root test later (after learning more about time-series analysis).
- We start with:

$$\begin{aligned} Y_t &= \rho Y_{t-1} + e_t \\ \rightarrow \quad Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + e_t \\ \rightarrow \quad \triangle Y_t &= (\rho - 1)\, Y_{t-1} + e_t \end{aligned}$$

  - Let's define $\delta = \rho - 1$, then:

$$\triangle Y_t = \delta Y_{t-1} + e_t$$

Unit Root Test [cont'd]

- We estimate the above equation, and test the null hypothesis $\delta = 0$ against the alternative $\delta < 0$.
  - If $\rho = 1$ (unit root), $\delta = 0$.
  - If $-1 < \rho < 1$, then δ is negative.
- If we reject the null hypothesis, then we can conclude the process is stationary.
- If we do not reject the null, then we should strongly suspect there is a unit root (or, the process is non-stationary).
- We will see an example of the practical unit root test in the Gretl session.
  - Augmented Dicky-Fuller test (ADF test)

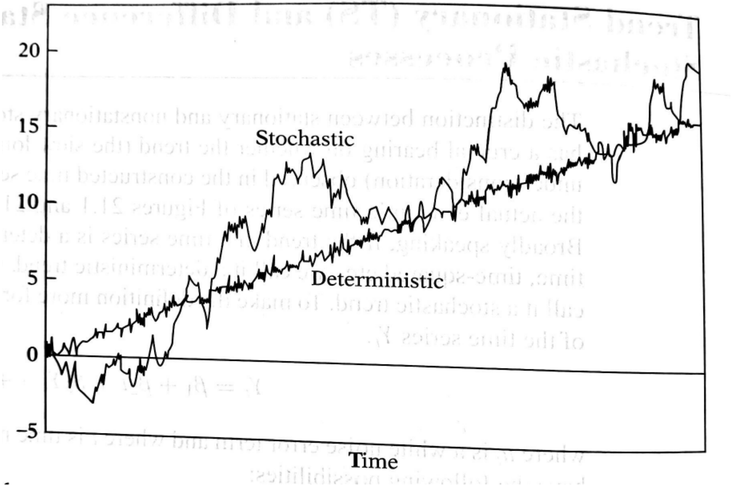
Difference Stationary and Trend Stationary Stochastic Process

Deterministic and Stochastic Trend

- As we discussed before, we are interested in a stationary process, but not a non-stationary process.
  - We can make a non-stationary stochastic process to a stationary process.
  - The changing procedure is related to whether the trend observed in the time series is *deterministic* or *stochastic*.
- **Deterministic trend**: if the trend in a time series is a deterministic function of time, such as t, t^2, etc., we call it a deterministic trend.
- **Stochastic trend**: if the trend is not predictable, we call it a stochastic trend.

Deterministic and Stochastic Trend [cont'd]

DGP: $Y_t = 0.5 + Y_{t-1} + e_t, \; Y_t = 0.5t + e_t$
*source: Gujarati Figure 21.5



$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + e_t$$

- **Pure random walk** ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 1$)

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

  - This process is a random walk without drift and is therefore non-stationary.
  - But note that, as $\triangle Y_t = e_t$, it becomes stationary.
  - Hence, a pure random walk is a **difference stationary process (DSP)**.

- **Random walk with drift** ($\beta_1 \neq 0$, $\beta_2 = 0$, $\beta_3 = 1$)

$$Y_t = \beta_1 + Y_{t-1} + e_t$$

  - This process is a random walk with drift and is therefore non-stationary.
  - If we write it as

$$\triangle Y_t = \beta_1 + e_t$$

    this means Y_t exhibit a positive ($\beta_1 > 0$) or negative ($\beta_1 < 0$) trend.
  - Such a trend is called a **stochastic trend**.
  - A random walk with drift is a **DSP** because the non-stationarity in Y_t can be eliminated by taking first differences.

DSP and TSP: Deterministic Trend

- **Deterministic trend** ($\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 = 0$)

$$Y_t = \beta_1 + \beta_2 t + e_t$$

  - This process is called a **trend stationary process (TSP)**.
  - Although the mean of Y_t is $\beta_1 + \beta_2 t$, which is not constant, its variance $(= \sigma^2)$ is constant.
  - Note that this process does not have a unit root, so this process is called as *"non-stationary process without a unit root"*.
  - Once the value of β_1 and β_2 are known, the mean can be forecast perfectly.
  - Therefore, if we subtract the mean of Y_t form Y_t, the resulting series will be stationary $\rightarrow$ *"trend stationary"*

- **Random walk with drift and deterministic trend** ($\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 = 1$)

$$Y_t = \beta_1 + \beta_2 t + Y_{t-1} + e_t$$

  - We write the above equation as

$$\triangle Y_t = \beta_1 + \beta_2 t + e_t$$

    this means Y_t should be differenced and we should remove the deterministic trend as well.

  - If $|\beta_3| < 1$, the process is just a TSP.

Underdifferencing and Overdifferencing

- It is very important to apply the right sort of stationarity transform to the data, if they are not stationary!
- If a non-stationary time series is DSP but we treat it as TSP, this is called **underdifferencing**.
- If a non-stationary time series is TSP but we treat it as DSP, this is called **overdifferencing**.
- If we confuse a TSP series with a DSP series or vice versa, our goal (to obtain the stationary process) would not be achieved.
  - *(cf)* It is known that most financial market prices (stock prices, interest rates, etc.) are non-stationary because of stochastic rather than deterministic trend.

Remark: TSP vs. DSP

- Both **TSP** and **DSP** are non-stationary process.
  - TSP: The mean trend is **deterministic**.
  - DSP: The mean trend is **stochastic**.
- The distinction between a deterministic and stochastic trend has important implications for the long-term behavior of a process.
  - Time series with a **deterministic trend** (so, **TSP**) always revert to the trend in the long run (**mean-reverting**). That is, the effect of shocks are eventually eliminated.
  - Time series with a **stochastic trend** (so, **DSP**) never recover from shocks to the system. That is, the effect of shocks are **permanent**.
- As a result, if the non-stationary process is TSP, it is not a big deal.
- In that sense, when we say a time series is a non-stationary process, it usually means that the process is DSP.

Integrated Stochastic Process

Integrated Stochastic Process

- Time series that can be made stationary by differencing ($i.e.$ DSP) are often called "***integrated process***".

- For example, a random walk is non-stationary process and its first difference is stationary.
  - When a time series has to be differenced **once** to make it stationary, we say that the series is **integrated of order 1**.
  - That is, the random walk without drift is integrated process of order 1.

- Similarly, if a time series has to be differenced **twice** ($i.e.$ take the first difference of the first difference) to make it stationary, we call such a time series **integrated of order 2**.
  - For example, if Y_t is integrated of order 2, $\triangle^2 Y_t = \triangle\triangle Y_t = \triangle Y_t - \triangle Y_{t-1}$ $= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$ will become stationary.
  - Note that $\triangle^2 Y_t \neq Y_t - Y_{t-2}$.

Integrated Stochastic Process [cont'd]

- In general, if a time series has to be differenced d **times** to make it stationary, that time series is said to be **integrated of order d**.
  - A time series Y_t integrated of order d is denoted as $Y_t \sim I(d)$.
- If a time series is stationary to begin with ($i.e.$ it does not require any differencing), it is said to be integrated of order zero, denoted by $Y_t \sim I(0)$
  - Therefore, we use the term "stationary time series" and "time series integrated of order zero" to mean the same thing.
- Most economic time series are generally $I(1)$.
  - That is, they generally become stationary only after taking their first differences.

Properties of Integrated Series

- If $X_t \sim I(d)$, then $Z_t = a + bX_t \sim I(d)$, where a and b are constants.
    - A linear combination of an $I(d)$ series is also $I(d)$.
    - If $X_t \sim I(0)$, then $Z_t = a + bX_t \sim I(0)$.
- If $X_t \sim I(0)$ and $Y_t \sim I(1)$, then $Z_t = X_t + Y_t \sim I(1)$.
    - A linear combination or sum of stationary and non-stationary time series is non-stationary.
- If $X_t \sim I(d_1)$ and $Y_t \sim I(d_2)$, then $Z_t = aX_t + bY_t \sim I(d_2)$, where $d_1 < d_2$.
- If $X_t \sim I(d)$ and $Y_t \sim I(d)$, then $Z_t = aX_t + bY_t \sim I(d^*)$.
    - d^* is generally equal to d.
    - But in some cases, $d^* < d \rightarrow$ **Cointegration**

- We must pay careful attention in combining two or more time series that are integrated of different order. .
- *(optional)* To see why this is important, consider the two-variable simple regression model: $Y_t = \beta_1 + \beta_2 X_t + e_t$
  - Under the classical assumptions, we know that

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2}$$

  - Suppose that Y_t is $I(0)$ and X_t is $I(1)$.
  - Since X_t is non-stationary, its variance $\left(E\left(X_t - \bar{X}\right)^2\right)$ will increase indefinitely.
  - Note that $\frac{1}{n}\sum\left(X_t - \bar{X}\right)^2$ is the sample variance of X_t, which is the denominator of $\hat{\beta}_2$.
  - Thus, huge denominator dominates the numerator, so $\hat{\beta}_2$ will converge to to zero in large samples.

Wold Representation

Wold Representation

- **Wold decomposition theorem**: In general, any ***stationary*** stochastic process can be expressed as the sum of ***deterministic*** component and ***stochastic*** component.

- **Wold representation**

$$Y_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots$$

  - **Stochastic component**: a linear combination of lags of a white noise process.
  - **Deterministic component**: uncorrelated with the stochastic component, and 100% predictable
    - Deterministic component need not necessarily be linear.
    - For example, it could be sine wave (nonlinear but 100% predictable).
    - However, we will focus on the simple constant deterministic component μ from now on.

Key Assumptions of Wold Representation

- **Assumption 1**

$$e_t \sim iidN\left(0, \sigma^2\right)$$

  - e_t means a shock or a prediction error.
  - By this assumption, $E\left(Y_t\right) = \mu$.

- **Assumption 2**

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty, \text{where } \psi_0 = 1$$

  - Note that the variance of Y_t is:

$$Var\left(Y_t\right) = \sigma^2 + \psi_1^2\sigma^2 + \psi_2^2\sigma^2 + \cdots = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$$

  - Therefore, $Var\left(Y_t\right)$ is finite (and time-invariant).

Wold Representation: Important Remark

- Wold representation is the **unique** linear representation!
  - In other words, $\psi_1, \psi_2, \psi_3, \cdots$ are unique.
- If we find the unique ψ's of the Wold representation, we can predict the impact of new (unexpected) shock on Y_t.
  - **Impulse-Response**

$$\frac{\partial Y_t}{\partial e_{t-1}} = \psi_1, \ \frac{\partial Y_t}{\partial e_{t-2}} = \psi_2, \ \frac{\partial Y_t}{\partial e_{t-3}} = \psi_3, \ \cdots$$

  - Under the stationary condition, we can forecast the effect of e_t on the future value of Y_t.

$$\frac{\partial Y_{t+1}}{\partial e_t} = \psi_1, \ \frac{\partial Y_{t+2}}{\partial e_t} = \psi_2, \ \frac{\partial Y_{t+3}}{\partial e_t} = \psi_3, \ \cdots$$

- Recall that $\sum_{i=0}^{\infty} \psi_i^2 < \infty \iff \lim_{i \to \infty} \psi_i = 0$
  - Therefore, we can say that the further a shock is, the weaker the effect of the shock is.

Challege Can we estimate the all parameters of the Wold representation?

- Parameters in Wold representation: μ, σ^2, ψ_1, ψ_2, ψ_3, $\cdots \to (\infty + 2)$ parameters
- Unfortunately, our time series observations are finite.
- Therefore, we need to approximate the Wold representation (or stationary process) into something!

$\Longrightarrow$ **ARMA model**