Lecture 1. Dynamic Nature of Relationships

Narek Ohanyan

American University of Armenia

February 5, 2024

Outline

Time Series Data

Dynamic Relationships

Autoregressive and Moving Average Models

Testing for Serial Correlation

Nature of time series data

The nature of the data available has an important bearing on the appropriate choice of an econometric model.

Two features of time-series data to consider:

1. Time-series observations on a given economic unit, observed over a number of time periods, are likely to be correlated
2. Time-series data have a natural ordering according to time

Stationary and nonstationary time series

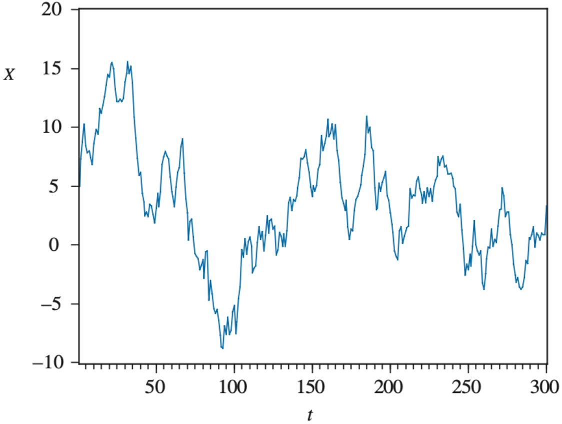
A time series is said to be **stationary** if its mean and variance are constant over time.

A time series is said to be **nonstationary** if its mean and variance are not constant over time.

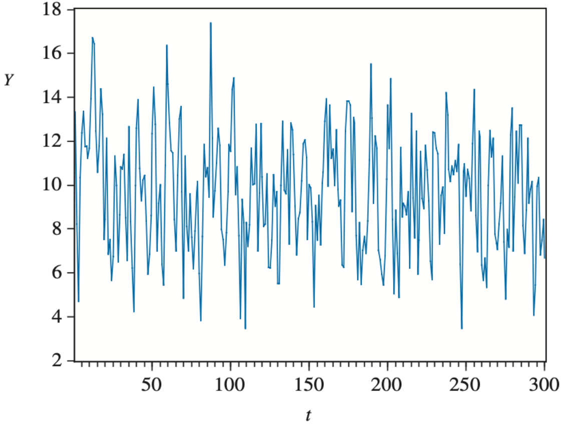
In other words, a stationary variable is one that is not trending, and nor wandering aimlessly without returning to its mean.

Stationarity is an important property because many time series models are based on the assumption that the data are stationary.

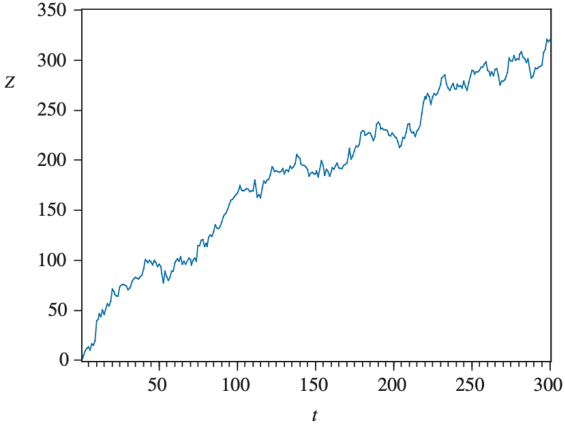
Different types of time series



Different types of time series



Different types of time series

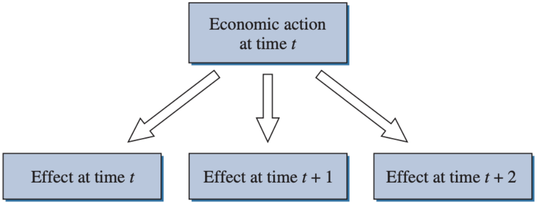


Outline

Dynamic relationships

In many economic applications, the relationship between two variables is **dynamic**.

That is, the change in a variable may have an impact on itself, or other variables, in one or more future time periods



Modelling dynamic relationships

There may be three different types of dynamic relationships:

1. The dependent variable y may be a function of current and past values of an explanatory variable x

$$y_t = f(x_t, x_{t-1}, x_{t-2}) + e_t$$

2. The dependent variable y may be a function of its own past values

$$y_t = f(y_{t-1}, x_t) + e_t$$

3. The error term u may be a function of its own past values

$$\begin{aligned} y_t &= f(x_t) + e_t \\ e_t &= g(e_{t-1}) + \varepsilon_t \end{aligned}$$

Distributed lag models

The first type of dynamic relationship is called a **distributed lag model**.

Suppose that the variable y depends on current and past values of variable x, up to q periods into the past

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t$$

- The coefficient β_0 is called the *impact* multiplier.
- The coefficients β_s are called the s-period *delay* multipliers.

Dynamic effects in distributed lag models

Recall that the distributed lag model is

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + e_t$$

Assume x_t is increased by one unit and then maintained at its new level in subsequent periods.

- The immediate impact at t will be β_0
- The effect in period $t+1$ will be $\beta_0 + \beta_1$
- The effect in period $t+2$ will be $\beta_0 + \beta_1 + \beta_2$ and so on

Th cumulative effect of a change in x_t on y_t is called **interim multiplier**.

The **total multiplier** is the final effect on y of the sustained increase after q or more periods have elapsed.

Autorergressive models

The second type of dynamic relationship is called an **autoregressive model**.

Suppose that the variable y depends on its own past values, up to p periods into the past

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

The model is called an autoregressive model of order p, or an $AR(p)$ model.

Dynamic effects in autoregressive models

Recall that the autoregressive model is

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

Suppose that y_t is increased by one unit and then maintained at its new level in subsequent periods.

- The effect in period $t+1$ will be ϕ_1
- The effect in period $t+2$ will be $\phi_1^2 + \phi_2$
- The effect in period $t+3$ will be $\phi_1^3 + 2\phi_1\phi_2 + \phi_3$ and so on

Autorergressive distributed lag models

We can combine the two types of dynamic relationships into a single model, called an **autoregressive distributed lag model**

$$y_t = \alpha + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + e_t$$

The model is called an autoregressive distributed lag model of order p and q, or an $ARDL(p, q)$ model.

Models with autoregressive errors

The third type of dynamic relationship is called a **model with autoregressive errors**.

Suppose that the error term e_t depends on its own past value

$$y_t = \alpha + \beta_0 x_t + e_t$$
$$e_t = \rho e_{t-1} + \varepsilon_t$$

Substituting the second equation into the first yields

$$y_t = \delta + \rho y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

where $\delta = \alpha (1 - \rho)$ and $\beta_1 = -\rho\beta_0$.

Hence, the model with autoregressive errors is a special case of an $ARDL(1, 1)$ model.

Outline

White noise and IID processes

The basic building block of time-series models is the *white noise* process.

A random variable e_t is called **white noise (WN)** if

$$\begin{aligned} &\mathrm{E}(e_t) = 0 \\ &\mathrm{Var}(e_t) = \sigma^2 \\ &\mathrm{Cov}(e_t, e_{t-s}) = 0 \quad \text{for all} \quad s \neq 0 \end{aligned}$$

A random variable e_t is called **independent and identically distributed (IID)** if

$$\begin{aligned} &F(e_t) = F(e_{t-s}) \quad \text{for all} \quad s \\ &e_t \perp e_{t-s} \quad \text{for all} \quad s \neq 0 \end{aligned}$$

where $F(e_t)$ is the cumulative distribution function of e_t.

An IID process is also a WN process.

Serial correlation

Autocorrelation, or serial correlation, is the correlation between a variable and its past or future values.

The **autocovariance** of a variable x at lag k is defined as

$$\mathrm{Cov}(x_t, x_{t-k}) = \mathrm{E}\left[(x_t - \mu_x)(x_{t-k} - \mu_x)\right]$$

Then, the **autocorrelation** coefficient of x at lag k is

$$\rho_k = \frac{\mathrm{Cov}(x_t, x_{t-k})}{\mathrm{Var}(x_t)}$$

The autocorrelation coefficients ρ_k take values between -1 and 1.

AR(1) model

The simplest autoregressive model is the $AR(1)$ model

$$y_t = c + \phi y_{t-1} + e_t \qquad e_t \sim WN(0, \sigma^2)$$

We assume that $|\phi| < 1$, so that the model is stationary.

The mean and variance of y_t are

$$\mathrm{E}(y_t) = \frac{c}{1-\phi} \qquad \mathrm{Var}(y_t) = \frac{\sigma^2}{1-\phi^2}$$

The autocovariance of y_t at lag k is

$$\mathrm{Cov}(y_t, y_{t-k}) = \phi^k \frac{\sigma^2}{1-\phi^2}$$

The autocorrelation of y_t at lag k is

$$\rho_k = \phi^k$$

MA(1) model

The simplest moving average model is the $MA(1)$ model

$$y_t = c + e_t + \theta e_{t-1} \qquad e_t \sim WN(0, \sigma^2)$$

The mean and variance of y_t are

$$\mathrm{E}(y_t) = c \qquad \mathrm{Var}(y_t) = \sigma^2 \left(1 + \theta^2\right)$$

The autocovariance of y_t at lag k is

$$\mathrm{Cov}(y_t, y_{t-k}) = \theta\sigma^2 \quad \text{for} \quad k = 1 \quad \text{and} \quad \mathrm{Cov}(y_t, y_{t-k}) = 0 \quad \text{for} \quad k \geq 2$$

The autocorrelation of y_t at lag k is

$$\rho_k = \frac{\theta}{1 + \theta^2} \quad \text{for} \quad k = 1 \quad \text{and} \quad \rho_k = 0 \quad \text{for} \quad k \geq 2$$

Autocorrelation and partial autocorrelation

Autocorrelation

Recall that the *autocorrelation* coefficient of y_t at lag k is the correlation between y_t and y_{t-k}

$$\rho_k = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\mathrm{Var}(y_t)}$$

Partial autocorrelation

The *partial autocorrelation* coefficient of y_t at lag k is the correlation between y_t and y_{t-k} after removing the effect of the intermediate lags $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$.

$$\phi_{k,k} = \frac{\mathrm{Cov}(y_t, y_{t-k} | y_{t-1}, \dots, y_{t-k+1})}{\mathrm{Var}(y_t | y_{t-1}, \dots, y_{t-k+1})}$$

ACF and PACF plots

Autocorrelation and partial autocorrelation functions plot the ρ_k and $\phi_{k,k}$ against k.

For an $AR(p)$ model

- The ACF plot will show gradual **decay** before and after lag p,
- The PACF plot will show a sharp **cutoff** after lag p.

For an $MA(q)$ model

- The ACF plot will show a sharp **cutoff** after lag q,
- The PACF plot will show gradual **decay** before and after lag q.

Hence, ACF and PACF plots can be used to identify the **order** of an AR or MA model.

Outline

Estimation and testing for autocorrelation

The sample autocorrelation coefficient of x at lag k is

$$\widehat{\rho}_k = \frac{\sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}$$

Since $\rho_k = 0$ means no autocorrelation, we are often interested in the following hypothesis $\mathrm{H}_0 : \rho_k = 0$ versus $\mathrm{H}_1 : \rho_k \neq 0$.

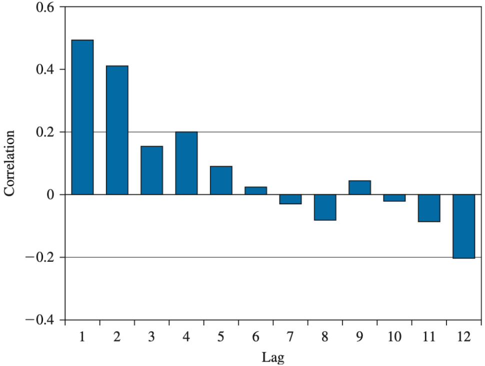
The test statistic is

$$\sqrt{T}\widehat{\rho}_k \sim \mathcal{N}(0, 1)$$

A sample autocorrelation function plot may also be used for visualizing and testing for autocorrelation.

Sample autocorrelation function (ACF)

A sample autocorrelation function (ACF) plot or **correlogram** is a plot of $\widehat{\rho}_k$ against k.



Confidence intervals of $\pm 1.96/\sqrt{T}$ are also commonly plotted to help identify significant autocorrelation coefficients.