Stationarity

Econ 211C – Unit 1, Section 1

Eric M. Aldrich

UC Santa Cruz

Time Series

A time series is a stochastic process indexed by time:

$$\{Y(t) : t \in \mathcal{T}\}.$$

Let's focus on the case when $Y(t)$ is univariate.

- If $\mathcal{T}$ is an interval in $\mathbb{R}$, then $Y(t)$ is a continuous time stochastic process.
- If $\mathcal{T}$ is a set of discrete indices, $Y(t)$ is a discrete time stochastic process.
  - In this case, we denote the time series process as $\{Y_t\}_{t\in\mathcal{T}}$ or simply $\{Y_t\}$.
  - Note that $\mathcal{T}$ could be an infinite set such as $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$: $\{Y_t\}_{t\in\mathcal{T}} = \{Y_t\}_{t=-\infty}^{\infty}$.
- In this course we will focus on discrete time series.

Distributions

We will think of $\{Y_t\}_{t\in\mathcal{T}}$ as a random variable in its own right.

- $\boldsymbol{y}_{\mathcal{T}} = \{y_t\}_{t\in\mathcal{T}}$ is a *single* realization of $\boldsymbol{Y}_{\mathcal{T}} = \{Y_t\}_{t\in\mathcal{T}}$.
- The CDF is $F_{\boldsymbol{Y}_{\mathcal{T}}}(\boldsymbol{y}_{\mathcal{T}})$ and the PDF is $f_{\boldsymbol{Y}_{\mathcal{T}}}(\boldsymbol{y}_{\mathcal{T}})$.
- For example, consider $\mathcal{T} = 1, \ldots, 100$:

$$F\left(\{y_t\}_{t=1}^{100}\right) = P(Y_1 \leq y_1, \ldots, Y_{100} \leq y_{100}).$$

- Notice that $\boldsymbol{Y}_{\mathcal{T}}$ is just a collection of random variables and $f_{\boldsymbol{Y}_{\mathcal{T}}}(\boldsymbol{y}_{\mathcal{T}})$ is the joint density.

Time Series Observations

As statisticians and econometricians, we want many observations of $\boldsymbol{Y}_{\mathcal{T}}$ to learn about its distribution:

$$\boldsymbol{y}_{\mathcal{T}}^{(1)}, \quad \boldsymbol{y}_{\mathcal{T}}^{(2)}, \quad \boldsymbol{y}_{\mathcal{T}}^{(3)}, \quad \ldots$$

Likewise, if we are only interested in the marginal distribution of Y_{17}

$$f_{Y_{17}}(a) = P(Y_{17} \leq a)$$

we want many observations: $\left\{ y_{17}^{(i)} \right\}_{i=1}^{N}$.

Time Series Observations

Unfortunately, we usually only have *one observation* of $\boldsymbol{Y}_{\mathcal{T}}$.

- Think of the daily closing price of Harley-Davidson stock since January 2nd.
- Think of your cardiogram for the past 100 seconds.

In neither case can you repeat history to observe a new sequence of prices or electronic heart signals.

- In time series econometrics we typically base inference on a single observation.
- Additional assumptions about the process will allow us to exploit information in the full sequence $\boldsymbol{y}_{\mathcal{T}}$ to make inferences about the joint distribution $F_{\boldsymbol{Y}_{\mathcal{T}}}(\boldsymbol{y}_{\mathcal{T}})$.

Moments

Since the stochastic process is comprised of individual random variables, we can consider moments of each:

$$
\begin{aligned}
\mathrm{E}\left[Y_t\right] &= \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t = \mu_t \\
\mathrm{Var}\left(Y_t\right) &= \int_{-\infty}^{\infty} (y_t - \mu_t)^2 f_{Y_t}(y_t) dy_t = \gamma_{0t} \\
\mathrm{Cov}\left(Y_t, Y_{t-j}\right) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y_t - \mu_t)(y_{t-j} - \mu_{t-j}) \\
&\qquad \times f_{Y_t, Y_{t-j}}(y_t, y_{t-j}) dy_t dy_{t-j} = \gamma_{jt},
\end{aligned}
$$

where f_{Y_t} and $f_{Y_t, Y_{t-j}}$ are the marginal distributions of $f_{\boldsymbol{Y}_{\mathcal{T}}}$ obtained by integrating over the appropriate elements of $\boldsymbol{Y}_{\mathcal{T}}$.

Autocovariance

Suppose $\boldsymbol{Y}_{\mathcal{T}} = (Y_1, Y_2, \ldots, Y_T)'$.

- $\left\{ \{\gamma_{jt}\}_{j=0}^{T-1} \right\}_{t=1}^{T}$ are the elements of the $T \times T$ covariance matrix of $\boldsymbol{Y}_{\mathcal{T}}$: γ_{jt} is the $(t, |j|+1)$ element of the matrix.

$$\Sigma_{\boldsymbol{Y}_{\mathcal{T}}} = \begin{bmatrix} \gamma_{0,1} & \gamma_{-1,1} & \cdots & \gamma_{-T+1,1} \\ \gamma_{1,2} & \gamma_{0,2} & \cdots & \gamma_{-T+2,2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1,T} & \gamma_{T-2,T} & \cdots & \gamma_{0,T} \end{bmatrix}$$

- This is a symmetric matrix.

- γ_{jt} is known as the jth autocovariance of Y_t since it is the covariance of Y_t with its own lagged value.

Autocorrelation

The jth autocorrelation of Y_t is defined as

$$\begin{aligned}\rho_{jt} &= Corr(Y_t, Y_{t-j}) \\ &= \frac{\operatorname{Cov}\left(Y_t, Y_{t-j}\right)}{\sqrt{\operatorname{Var}\left(Y_t\right)}\sqrt{\operatorname{Var}\left(Y_{t-j}\right)}} \\ &= \frac{\gamma_{jt}}{\sqrt{\gamma_{0t}}\sqrt{\gamma_{0t-j}}}\end{aligned}$$

Sample Moments

If we had N observations $\boldsymbol{y}_{\mathcal{T}}^{(1)}, \ldots, \boldsymbol{y}_{\mathcal{T}}^{(N)}$, we could estimate moments of each (univariate) Y_t in the usual way:

$$\hat{\mu}_t = \frac{1}{N} \sum_{i=1}^{N} y_t^{(i)}.$$

$$\hat{\gamma}_{0t} = \frac{1}{N} \sum_{i=1}^{N} (y_t^{(i)} - \hat{\mu}_t)^2.$$

$$\hat{\gamma}_{jt} = \frac{1}{N} \sum_{i=1}^{N} (y_t^{(i)} - \hat{\mu}_t)(y_{t-j}^{(i)} - \hat{\mu}_{t-j}).$$

Example

Suppose $\boldsymbol{Y}_{\mathcal{T}}$ is a T dimensional vector with each element described by

$$Y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2), \forall t.$$

We could express this in vector form

$$\boldsymbol{Y}_{\mathcal{T}} = \boldsymbol{\mu}_{\mathcal{T}} + \boldsymbol{\varepsilon}_{\mathcal{T}}$$

where

$$\boldsymbol{\varepsilon}_{\mathcal{T}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}_{T \times T},$$

and where $\boldsymbol{\mu}_{\mathcal{T}} = (\mu_1, \mu_2, \ldots, \mu_T)'$ and $\mathbf{0} = (0, 0, \ldots, 0)'_{1 \times T}$.

Example

In this case,

$$\begin{aligned}
\mu_t &= \mathrm{E}\left[Y_t\right] = \mu_t, \;\; \forall t, \\
\gamma_{0t} &= \mathrm{Var}\left(Y_t\right) = \mathrm{Var}\left(\varepsilon_t\right) = \sigma_t^2, \;\; \forall t \\
\gamma_{jt} &= \mathrm{Cov}\left(Y_t, Y_{t-j}\right) = \mathrm{Cov}\left(\varepsilon_t, \varepsilon_{t-j}\right) = 0, \;\; \forall t, j \neq 0.
\end{aligned}$$

- If $\sigma_t^2 = \sigma^2 \; \forall t$, $\boldsymbol{\varepsilon}_{\mathcal{T}}$ is known as a *Gaussian white noise* process.
- In this case, $\boldsymbol{Y}_{\mathcal{T}}$ is a Gaussian white noise process with drift – $\boldsymbol{\mu}_{\mathcal{T}}$ is the drift vector.

White Noise

Generally speaking, $\boldsymbol{\varepsilon}_{\mathcal{T}}$ is a *white noise process* if

$$\mathrm{E}\left[\varepsilon_t\right] = 0, \ \ \forall t \tag{1a}$$

$$\mathrm{E}\left[\varepsilon_t^2\right] = \sigma^2, \ \ t \tag{1b}$$

$$\mathrm{E}\left[\varepsilon_t \varepsilon_\tau\right] = 0, \quad \text{for } t \neq \tau. \tag{1c}$$

Notice there is no distributional assumption for ε_t.

- If ε_t and ε_τ are independent for $t \neq \tau$, $\boldsymbol{\varepsilon}_{\mathcal{T}}$ is *independent white noise.*
- Notice that independence $\Rightarrow$ Equation (1c), but Equation (1c) $\not\Rightarrow$ independence.
- If $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ $\forall t$, as in the example above, $\boldsymbol{\varepsilon}_{\mathcal{T}}$ is Gaussian white noise.

Weak Stationarity

Suppose the first and second moments of a stochastic process $\boldsymbol{Y}_{\mathcal{T}}$ don't depend on $t \in \mathcal{T}$:

$$\mathrm{E}\left[Y_t\right] = \mu \quad \forall t$$
$$\mathrm{Cov}\left(Y_t, Y_{t-j}\right) = \gamma_j \quad \forall t \text{ and any } j.$$

- In this case $\boldsymbol{Y}_{\mathcal{T}}$ is *weakly stationary* or *covariance stationary*.
- In the previous example, if $Y_t = \mu + \varepsilon_t \; \forall t$, $\boldsymbol{Y}_{\mathcal{T}}$ is weakly stationary.
- However if $\mu_t \neq \mu \; \forall t$, $\boldsymbol{Y}_{\mathcal{T}}$ is *not* weakly stationary.

Autocorrelation under Weak Stationarity

If $\boldsymbol{Y}_{\mathcal{T}}$ is weakly stationary

$$\begin{aligned}
\rho_{jt} &= \frac{\gamma_{jt}}{\sqrt{\gamma_{0t}}\sqrt{\gamma_{0t-j}}} \\
&= \frac{\gamma_j}{\sqrt{\gamma_0}\sqrt{\gamma_0}} \\
&= \frac{\gamma_j}{\gamma_0} \\
&= \rho_j.
\end{aligned}$$

- Note that $\rho_0 = 1$.

Weak Stationarity

Under weak stationarity, autocovariances γ_j only depend on the distance between random variables within a stochastic process:

$$\text{Cov}\left(Y_\tau, Y_{\tau-j}\right) = \text{Cov}\left(Y_t, Y_{t-j}\right) = \gamma_j.$$

This implies

$$\gamma_{-j} = \text{Cov}\left(Y_{t+j}, Y_t\right) = \text{Cov}\left(Y_t, Y_{t-j}\right) = \gamma_j.$$

More generally,

$$\Sigma_{\boldsymbol{Y}_\mathcal{T}} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} & \gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{T-2} & \gamma_{T-3} & \cdots & \gamma_0 & \gamma_1 \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}.$$

Strict Stationarity

$\boldsymbol{Y}_{\mathcal{T}}$ is *strictly stationary* if for any set $\{j_1, j_2, \ldots, j_n\} \in \mathcal{T}$

$$f_{Y_{j_1},\ldots,Y_{j_n n}}(a_1, \ldots, a_n) = f_{Y_{j_1+\tau},\ldots,Y_{j_n n+\tau}}(a_1, \ldots, a_n), \;\; \forall\tau.$$

- Strict stationarity means that the joint distribution of any subset of random variables in $\boldsymbol{Y}_{\mathcal{T}}$ is invariant to shifts in time, τ.
- Strict stationarity $\Rightarrow$ weak stationarity if the first and second moments of a stochastic process exist.
- Weak stationarity $\not\Rightarrow$ strict stationarity: invariance of first and second moments to time shifts (weak stationarity) does not mean that all higher moments are invariant to time shifts (strict stationarity).

Strict Stationarity

If $\boldsymbol{Y}_{\mathcal{T}}$ is Gaussian then weak stationarity $\Rightarrow$ strict stationarity.

- If $\boldsymbol{Y}_{\mathcal{T}}$ is Gaussian, all marginal distributions of $(Y_{j_1}, \ldots, Y_{j_n})$ are also Gaussian.
- Gaussian distributions are fully characterized by their first and second moments.

Ergodicity

Given N identically distributed weakly stationary stochastic processes $\{\boldsymbol{Y}_{\mathcal{T}}\}_{i=1}^{N}$, the *ensemble average*

$$\frac{1}{N}\sum_{i=1}^{N} Y_t^{(i)} \xrightarrow{p} \mu, \quad \forall t \in \mathcal{T}.$$

For a single stochastic process, we desire conditions under which the *time average*

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{p} \mu, \tag{2}$$

where we have assumed $\mathcal{T} = \{1, \ldots, T\}$.

Ergodicity

If $\boldsymbol{Y}_{\mathcal{T}}$ is weakly stationary and

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty, \tag{3}$$

$\boldsymbol{Y}_{\mathcal{T}}$ is *ergodic for the mean* and Equation (2) holds.

- Equation (3) requires that the autocovariances fall to zero sufficiently quickly.
- i.e. a *long realization* of $\{y_t\}$ will have many segments that are uncorrelated and which can be used to approximate an ensemble average.

Ergodicity

A weakly stationary process is ergodic for the second moments if

$$\frac{1}{T-j}\sum_{t=j+1}^{T}(Y_t-\mu)(Y_{t-j}-\mu)\xrightarrow{p}\gamma_j. \tag{4}$$

- Separate conditions exist which cause Equation (4) to hold.
- If $\boldsymbol{Y}_{\mathcal{T}}$ is Gaussian and stationary, then Equation (3) ensures that $\boldsymbol{Y}_{\mathcal{T}}$ is ergodic for all moments.