Definitions and Stationarity

10

Similarly to the univariate case, we start our exposition with the concept of stationarity which is also crucial in the multivariate setting. Before doing so let us define the multivariate stochastic process.

Definition 10.1. *A multivariate stochastic Process, $\{X_t\}$, is a family of random variables indexed by t, $t \in \mathbb{Z}$, which take values in $\mathbb{R}^n$, $n \geq 1$. n is called the dimension of the process.*

Setting $n = 1$, the above definition includes as a special case univariate stochastic processes. This implies that the statements for multivariate processes carry over analogously to the univariate case. We view X_t as a column vector:

$$X_t = \begin{pmatrix} X_{1t} \\ \vdots \\ X_{nt} \end{pmatrix}.$$

Each element $\{X_{it}\}$ thereby represents a particular variable which may be treated as a univariate process. As in the example of Sect. 15.4.5, $\{X_t\}$ represents the multivariate process consisting of the growth rate of GDP Y_t, the unemployment rate U_t, the inflation rate P_t, the wage inflation rate W_t, and the growth rate of money M_t. Thus, $X_t = (Y_t, U_t, P_t, W_t, M_t)'$.

As in the univariate case, we characterize the joint distribution of the elements X_{it} and X_{jt} by the first two moments (if they exist), i.e. by the mean and the variance, respectively covariance:

$$\begin{aligned} \mu_{it} &= \mathbb{E}X_{it}, \qquad i = 1, \ldots, n \\ \gamma_{ij}(t, s) &= \mathbb{E}(X_{it} - \mu_{it})(X_{js} - \mu_{js}), \qquad i, j = 1, \ldots, n; t, s \in \mathbb{Z} \end{aligned} \tag{10.1}$$

It is convenient to write these entities compactly as vectors and matrices:

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_10

$$\mu_t = \begin{pmatrix} \mu_{1t} \\ \vdots \\ \mu_{nt} \end{pmatrix} = \mathbb{E}X_t = \begin{pmatrix} \mathbb{E}X_{1t} \\ \vdots \\ \mathbb{E}X_{nt} \end{pmatrix}$$

$$\Gamma(t,s) = \begin{pmatrix} \gamma_{11}(t,s) & \dots & \gamma_{1n}(t,s) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(t,s) & \dots & \gamma_{nn}(t,s) \end{pmatrix} = \mathbb{E}(X_t - \mu_t)(X_s - \mu_s)'$$

Thus, we apply the expectations operator element-wise to vectors and matrices. The matrix-valued function $\Gamma(t,s)$ is called the *covariance function* of $\{X_t\}$.

In analogy to the univariate case, we define stationarity as the invariance of the first two moments to time shifts:

Definition 10.2 (Stationarity)**.** *A multivariate stochastic process* $\{X_t\}$ *is* stationary *if and only if for all integers r, s and t we have:*

(i) $\mu = \mu_t = \mathbb{E}X_t$ *is constant (independent of t);*
(ii) $\mathbb{E}X_t'X_t < \infty$*;*
(iii) $\Gamma(t,s) = \Gamma(t+r, s+r)$.

In the literature these properties are often called weak stationarity, covariance stationarity, or stationarity of second order. If $\{X_t\}$ is stationary, the covariance function only depends on the number of periods between t and s (i.e. on $t-s$) and not on t or s themselves. This implies that by setting $r = -s$ and $h = t - s$ the covariance function simplifies to

$$\begin{aligned} \Gamma(h) = \Gamma(t-s) &= \Gamma(t+r, s+r) = \Gamma(t+h, t) \\ &= \mathbb{E}(X_{t+h} - \mu)(X_t - \mu)' = \mathbb{E}(X_t - \mu)(X_{t-h} - \mu)'. \end{aligned}$$

For $h = 0$, $\Gamma(0)$ is the unconditional covariance matrix of X_t. Using the definition of the covariances in Eq. (10.1) we get:

$$\Gamma(h) = \Gamma(-h)'.$$

Note that $\Gamma(h)$ is in general not symmetric for $h \neq 0$ because $\gamma_{ij}(h) \neq \gamma_{ji}(h)$ for $h \neq 0$.

Based on the covariance function of a stationary process, we can define the *correlation function* $R(h)$ where $R(h) = (\rho_{ij}(h))_{i,j}$ with

$$\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}.$$

In the case $i \neq j$ we refer to the *cross-correlations* between two variables $\{X_{it}\}$ and $\{X_{jt}\}$. The correlation function can be written in matrix notation as

$$R(h) = V^{-1/2}\Gamma(h)V^{-1/2}$$

where V represents a diagonal matrix with diagonal elements equal to $\gamma_{ii}(0)$. Clearly, $\rho_{ii}(0) = 1$. As for the covariance matrix we have that in general $\rho_{ij}(h) \neq \rho_{ji}(h)$ for $h \neq 0$. It is possible that $\rho_{ij}(h) > \rho_{ij}(0)$. We can summarize the properties of the covariance function by the following theorem.[1]

Theorem 10.1. *The covariance function of a stationary process $\{X_t\}$ has the following properties:*

(i) *For all $h \in \mathbb{Z}$, $\Gamma(h) = \Gamma(-h)'$;*
(ii) *for all $h \in \mathbb{Z}$, $|\gamma_{ij}(h)| \leq \sqrt{\gamma_{ii}(0) \times \gamma_{jj}(0)}$;*
(iii) *for each $i = 1, \ldots, n$, $\gamma_{ii}(h)$ is a univariate autocovariance function;*
(iv) $\sum_{r,k=1}^{m} a_r'\Gamma(r-k)a_k \geq 0$ *for all $m \in \mathbb{N}$ and all $a_1, \ldots, a_m \in \mathbb{R}^n$. This property is called non-negative definiteness (see Property 4 in Theorem 1.1 of Sect. 1.3 in the univariate case).*

Proof. Property (i) follows immediately from the definition. Property (ii) follows from the fact that the correlation coefficient is always smaller or equal to one. $\gamma_{ii}(h)$ is the autocovariance function of $\{X_{it}\}$ which delivers property (iii). Property (iv) follows from $\mathbb{E}\left(\sum_{k=1}^{m} a_k'(X_{t-k} - \mu)\right)^2 \geq 0$. □

If not only the first two moments are invariant to time shifts, but the distribution as a whole we arrive at the concept of strict stationarity.

Definition 10.3 (Strict Stationarity)**.** *A process multivariate $\{X_t\}$ is called* strictly stationary *if and only if, for all $n \in \mathbb{N}$, $t_1, \ldots, t_n$, $h \in \mathbb{Z}$, the joint distributions of $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+h}, \ldots, X_{t_n+h})$ are the same.*

An Example

Consider the following example for $n = 2$:

$$X_{1t} = Z_t$$
$$X_{2t} = Z_t + 0.75Z_{t-2}$$

where $Z_t \sim \mathrm{WN}(0, 1)$. We then have $\mu = \mathrm{E}X_t = 0$. The covariance function is given by

[1] We leave it to the reader to derive an analogous theorem for the correlation function.

$$\Gamma(h) = \begin{cases} \begin{pmatrix} 1 & 1 \\ 1 & 1.5625 \end{pmatrix}, & h = 0; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & h = 1; \\ \begin{pmatrix} 0 & 0 \\ 0.75 & 0.75 \end{pmatrix}, & h = 2. \end{cases}$$

The covariance function is zero for $h > 2$. The values for $h < 0$ are obtained from property (i) in Theorem 10.1. The correlation function is:

$$R(h) = \begin{cases} \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}, & h = 0; \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & h = 1; \\ \begin{pmatrix} 0 & 0 \\ 0.60 & 0.48 \end{pmatrix}, & h = 2. \end{cases}$$

The correlation function is zero for $h > 2$. The values for $h < 0$ are obtained from property (i) in Theorem 10.1.

One idea in time series analysis is to construct more complicated process from simple ones, for example by taking moving-averages. The simplest process is the white noise process which is uncorrelated with its own past. In the multivariate context the white noise process is defined as follows.

Definition 10.4. *A stochastic process $\{Z_t\}$ is called (multivariate)* white noise *process with mean zero and covariance matrix $\Sigma > 0$, denoted by $Z_t \sim \mathrm{WN}(0, \Sigma)$, if it is stationary $\{Z_t\}$ and*

$$\mathbb{E}Z_t = 0,$$

$$\Gamma(h) = \begin{cases} \Sigma, & h = 0; \\ 0, & h \neq 0. \end{cases}$$

If $\{Z_t\}$ is not only white noise, but independently and identically distributed we write $Z_t \sim \mathrm{IID}(0, \Sigma)$.

Remark 10.1. Even if each component of $\{Z_{it}\}$ is univariate white noise, this does not imply that $\{Z_t\} = \{(Z_{1t}, \ldots, Z_{nt})'\}$ is multivariate white noise. Take, for example the process $Z_t = (u_t, u_{t-1})'$ where $u_t \sim \mathrm{WN}(0, \sigma_u^2)$. Then $\Gamma(1) = \begin{pmatrix} 0 & 0 \\ \sigma_u^2 & 0 \end{pmatrix} \neq 0$.

Taking moving averages of a white noise process it is possible to generate new stationary processes. This leads to the definition of a linear process.

Definition 10.5. *A stochastic process $\{X_t\}$ is called* linear *if there exists a representation*

$$X_t = \sum_{j=-\infty}^{\infty} \Psi_j Z_{t-j}$$

where $Z_t \sim \text{IID}(0, \Sigma)$ *and where the sequence* $\{\Psi_j\}$ *of the* $n \times n$ *matrices is absolutely summable, i.e.* $\sum_{j=-\infty}^{\infty} \|\Psi_j\| < \infty$. *If for all* $j < 0$ $\Psi_j = 0$, *the linear process is also called an* $\text{MA}(\infty)$ *process.*

Theorem 10.2. *A linear process is stationary with a mean of zero and with covariance function*

$$\Gamma(h) = \sum_{j=-\infty}^{\infty} \Psi_{j+h} \Sigma \Psi_j' = \sum_{j=-\infty}^{\infty} \Psi_j \Sigma \Psi_{j-h}', \qquad h = 0, \pm 1, \pm 2, \ldots$$

Proof. The required result is obtained by applying the properties of $\{Z_t\}$ to $\Gamma(h) = \mathbb{E} X_{t+h} X_t' = \lim_{m \to \infty} \mathbb{E}\left(\sum_{j=-m}^{m} \Psi_j Z_{t+h-j}\right)\left(\sum_{k=-m}^{m} \Psi_k Z_{t-k}\right)'$. □

Remark 10.1. The same conclusion is reached if $\{Z_t\}$ is a white noise process and not an IID process.

Appendix: Norm and Summability of Matrices

As in the definition of a linear process it is often necessary to analyze the convergence of a sequence of matrices $\{\Psi_j\}$, $j = 0, 1, 2, \ldots$. For this we need to define a norm for matrices. The literature considers different alternative approaches. For our purposes, the choice is not relevant as all norms are equivalent in the finite dimensional vector space. We therefore choose the Frobenius, Hilbert-Schmidt or Schur norm which is easy to compute.[2] This norm treats the elements of a $m \times n$ matrix $A = (a_{ij})$ as an element of the $\mathbb{R}^{m \times n}$ Euclidean space and defines the length of A, denoted by $\|A\|$, as $\|A\| = \sqrt{\sum_{i,j} a_{ij}^2}$. This leads to the formal definition below.

Definition 10.6. *The Frobenius, Hilbert-Schmidt or Schur norm of a* $m \times n$ *matrix* A, *denoted by* $\|A\|$, *is defined as:*

$$\|A\|^2 = \sum_{i,j} a_{ij}^2 = \text{tr}(A'A) = \sum_{i=1}^{n} \lambda_i$$

where $\text{tr}(A'A)$ *denotes the trace of* $A'A$, *i.e. the sum of the diagonal elements of* $A'A$, *and where* λ_i *are the* n *eigenvalues of* $A'A$.

[2] For details see Meyer (2000, 279ff).

The matrix norm has the following properties:

$$\|A\| \geq 0 \quad \text{and} \quad \|A\| = 0 \text{ is equivalent to } A = 0,$$

$$\|\alpha A\| = |\alpha|\|A\| \text{ for all } \alpha \in \mathbb{R},$$

$$\|A\| = \|A'\|,$$

$$\|A + B\| \leq \|A\| + \|B\| \text{ for all matrices } A \text{ and } B \text{ of the same dimension},$$

$$\|AB\| \leq \|A\|\|B\| \text{ for all conformable matrices } A \text{ and } B.$$

The last property is called *submultiplicativity*.

A sequence of matrices $\{\Psi_j\}, j = 0, 1, \ldots$, is called *absolutely summable* if and only if $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$; the sequence is said to be *quadratic summable* if and only if $\sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty$. The absolute summability implies the quadratic summability, but not vice versa.

Corollary 10.1. *Absolute summability of* $\{\Psi_j\}$ *is equivalent to the absolute summability of each sequence* $\{[\Psi_j]_{kl}\}$, $k, l = 1, \ldots, n$, *i.e. to* $\lim_{j\to\infty} |[\Psi_j]_{kl}|$ *exists and is finite.*

Proof. In particular, he have:

$$|[\Psi_j]_{kl}| \leq \|\Psi_j\| = \sqrt{\sum_{k=1}^{n}\sum_{l=1}^{n}[\Psi_j]_{kl}^2} \leq \sum_{k=1}^{n}\sum_{l=1}^{n}|[\Psi_j]_{kl}|.$$

Summation over j gives

$$\sum_{j=0}^{\infty}|[\Psi_j]_{kl}| \leq \sum_{j=0}^{\infty}\|\Psi_j\| \leq \sum_{j=0}^{\infty}\sum_{k=1}^{n}\sum_{l=1}^{n}|[\Psi_j]_{kl}| = \sum_{k=1}^{\infty}\sum_{l=1}^{n}\sum_{j=0}^{n}|[\Psi_j]_{kl}|$$

The absolute convergence of each sequence $\{[\Psi_j]_{kl}\}$, $k, l = 1, \ldots, n$, follows from the absolute convergence of $\{\Psi_j\}$ by the first inequality. Conversely, the absolute convergence of $\{\Psi_j\}$ is implied by the absolute convergence of each sequence $\{[\Psi_j]_{kl}\}$, $k, l = 1, \ldots, n$, from the second inequality. □