Chapter 2

Linear Time Series Models

2.1 Linear Time Series Models

Prediction can be improved with better understanding of the stationary process. Very often, the stationary component is not $WN(0, \sigma^2)$; there are correlations. The next task is to build a suitable family of models for the stationary process. The classical models for the stationary process are *linear processes*.

Definition 2.1 (Linear process, Strictly Linear Process)**.** *A process* $\{X_t, t \in \mathbb{Z}\}$ *is said to be a* linear process *if it has representation*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad \{\epsilon_t\} \sim WN(0, \sigma^2), \tag{2.1}$$

where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$*. A stationary time series* $\{X_t\}$ *is* strictly linear *if it has the representation*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \qquad \{\epsilon_t\} \sim IID(0, \sigma^2).$$

where $\sum_j |\psi_j| < +\infty$*.*

This representation is taken to mean the following: let

$$X_{tm} = \mu + \sum_{j=-m}^{m} \psi_j \epsilon_{t-j} \qquad \{\epsilon_t\} \sim IID(0, \sigma^2) \quad \text{or} \quad \{\epsilon_t\} \sim WN(0, \sigma^2).$$

Then for each t,

$$\lim_{m_1 \to +\infty} \sup_{m_2 \geq m_1} \mathbb{E}\left[|X_{tm_1} - X_{tm_2}|^2\right] = 0.$$

The following lemma gives conditions under which an infinite sum may be approximated by a finite sum.

Lemma 2.2. *Let* $\{X_t\}$ *be sequence of random variables (possibly complex valued) that satisfies*

$$\sup_t \mathbb{E}[|X_t|^2] < \infty.$$

If

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

then the sequence

$$Y_{t,n} = \sum_{j=-n}^{n} \psi_j X_{t-j}$$

is Cauchy in L^2, in the sense that

$$\lim_{n\to+\infty} \sup_{m\geq 0} \mathbb{E}[(Y_{t,n+m} - Y_{t,n})^2] = 0.$$

Proof Suppose that $\sup_t \mathbb{E}[|X_t|^2] < \infty$. Then

$$\begin{aligned}
\mathbb{E}\left[|Y_{t,n+m} - Y_{t,n}|\right] &= \mathbb{E}\left[\left|\sum_{|j|=n+1}^{n+m} \psi_j X_{t-j}\right|^2\right] \\
&= \sum_{|j|=n+1}^{n+m} \sum_{|k|=n+1}^{n+m} \psi_j \psi_k \mathbb{E}\left[X_{t-j}\overline{X}_{t-k}\right] \leq \sup_t \mathbb{E}\left[|X_t|^2\right] \left(\sum_{|j|=n+1}^{n+m} |\psi_j|\right)^2
\end{aligned}$$

from which the result follows, since

$$\lim_{n\to+\infty} \sup_m \sum_{|j|=n+1}^{n+m} |\psi_j| \leq \lim_{n\to+\infty} \sum_{|j|=n+1}^{\infty} |\psi_j| = 0.$$

□

2.1.1 The Spectral Density

The *spectral density* of a stationary process is defined as follows:

Definition 2.3 (Spectral Density)**.** *Let γ be the ACVF for a stationary time series. The function f defined by*

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda}\gamma(h), \quad -\pi \leq \lambda \leq \pi, \tag{2.2}$$

is the spectral density *of γ. It is well defined if $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.*

The ACVF may be recovered from the spectral density:

$$\int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda = \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{i(h-k)\lambda}\gamma(k)\, d\lambda = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \int_{-\pi}^{\pi} e^{i(h-k)\lambda}\, d\lambda = \gamma(h).$$

The spectral density satisfies (among other things):

$$f(0) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h).$$

Properties of Autocovariance and Spectral Density The following result gives the autocovariance and spectral density in terms of ψ and σ^2.

Theorem 2.4. *A linear process* $\{X_t,\, t \in \mathbb{Z}\}$ *with representation given by* (2.1) *is stationary with mean* μ*, autocovariance function*

$$\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}, \tag{2.3}$$

and spectral density

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2, \tag{2.4}$$

where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$.

Proof Taking expectations,

$$\mathbb{E}[X_t] = \mathbb{E}\left[\mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}\right] = \mu + \sum_{j=-\infty}^{\infty} \psi_j \mathbb{E}[\epsilon_{t-j}] = \mu$$

and

$$\begin{aligned}
\gamma_X(h) &= \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)] = \mathbb{E}\left[\Big(\sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t+h-j}\Big)\Big(\sum_{k=-\infty}^{\infty} \psi_k \epsilon_{t-k}\Big)\right] \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \mathbb{E}[\epsilon_{t+h-j} \epsilon_{t-k}] \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \mathbb{E}[\epsilon_{h+k-j} \epsilon_0] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_{h+j} \psi_k \mathbb{E}[\epsilon_{k-j} \epsilon_0] \\
&= \sum_{j=-\infty}^{\infty} \psi_{h+j} \psi_j \mathbb{E}[\epsilon_0 \epsilon_0],
\end{aligned}$$

and (2.3) follows since $\mathbb{E}[\epsilon_0 \epsilon_0] = \sigma^2$.

Equation (2.2) now gives:

$$\begin{aligned}
f_X(\lambda) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma_h(h) = \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \\
&= \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} e^{-ih\lambda} \psi_j \psi_{j+h} = \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} e^{ij\lambda} \psi_j e^{-i(j+h)\lambda} \psi_{j+h} \\
&= \frac{\sigma^2}{2\pi} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} e^{ij\lambda} \psi_j e^{-ik\lambda} \psi_k = \frac{\sigma^2}{2\pi} \psi(e^{i\lambda}) \psi(e^{-i\lambda}) = \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2.
\end{aligned}$$

□

2.2 Prediction of random variables

Here we consider the problem of predicting a random variable Y from arbitrary random variables $W_1, \ldots, W_n$. The application will be to *time series* where we aim to predict X_{n+1} from $X_n, \ldots, X_1$. Since we expect the most recent observations to be more important for prediction, it is useful to consider a reverse order of the indices.

Consider any random variables $W_1, W_2, \ldots, W_n$ and Y with finite means and variances. Set $\mu_i = \mathbb{E}[W_i]$ and $\mu = \mathbb{E}[Y]$ and set $\Gamma_{n:ij} = \mathrm{Cov}(W_{n-i+1}, W_{n-j+1})$. Let Γ_n denote the matrix with entries $\Gamma_{n:ij}$. Let $\underline{\gamma} = (\gamma_1, \ldots, \gamma_n)^t$, where $\gamma_j = \mathrm{Cov}(W_{n-j+1}, Y)$.
The aim is to establish a linear predictor $P_{\mathcal{F}^{(W)}_{1:n}}(Y)$ of Y in terms of $W_1, W_2, \ldots, W_n$; that is, a random variable $P_{\mathcal{F}^{(W)}_{1:n}}(Y)$ of the form

$$P_{\mathcal{F}^{(W)}_{1:n}}(Y) = a_0 + a_1 W_n + \ldots + a_n W_1. \tag{2.5}$$

Let $\widehat{Y} = P_{\mathcal{F}^{(W)}_{1:n}}(Y)$ denote the predictor of Y. For convenience, and without loss of generality (since the means are assumed to be known), this may be centred:

$$\widehat{Y} - \mu = \tilde{a}_0 + a_1(W_n - \mu_n) + \ldots + a_n(W_1 - \mu_1).$$

Here $a_0 = \tilde{a}_0 + \mu - \sum_{j=1}^n a_j \mu_{n-j}$; the problems are equivalent.
Set $S(\tilde{a}_0, a_1, \ldots, a_n) = \mathbb{E}\left[(Y - \widehat{Y})^2\right]$ and choose $\tilde{a}_0, a_1, \ldots, a_n$ that minimise $S(\tilde{a}_0, a_1, \ldots, a_n)$. It follows from the definition that

$$\begin{aligned} S(\tilde{a}_0, a_1, \ldots, a_n) &= \mathbb{E}\left[((Y-\mu) - \tilde{a}_0 - a_1(W_n - \mu_n) - \ldots - a_n(W_1 - \mu_1))^2\right] \\ &= \tilde{a}_0^2 + \mathbb{E}\left[((Y-\mu) - a_1(W_n - \mu_n) - \ldots - a_n(W_1 - \mu_1))^2\right]. \end{aligned}$$

From this, it follows directly that $\tilde{a}_0 = 0$ (and hence, in the problem with arbitrary means, $a_0 = \mu - \sum_{j=1}^n a_j \mu_{n-j}$). For the others,

$$\begin{aligned} \frac{\partial S}{\partial a_i} &= -2\mathbb{E}\left[(W_{n-i+1} - \mu_{n-i+1})\left((Y-\mu) - a_1(W_n - \mu_n) - \ldots - a_n(W_1 - \mu_1)\right)\right] \\ &= -2\left(\gamma_i - \sum_{j=1}^n a_j \Gamma_{n;i,j}\right), \quad i = 1, \ldots, n. \end{aligned}$$

Since S is quadratic in $(a_0, \ldots, a_n)$ and goes to infinity as each $|a_j| \to +\infty$, it follows that the minimum is attained when all the partial derivatives are equal to zero;

$$\frac{\partial S}{\partial a_i} = 0 \quad \Leftrightarrow \quad \gamma_i = \sum_{j=1}^n a_j \Gamma_{n;i,j}, \quad i = 1, \ldots, n. \tag{2.6}$$

Set

$$\underline{a}_n = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Equation (2.6) may then be rewritten as:

$$\underline{\gamma}_n = \Gamma_n \underline{a}_n \quad \text{or, if } \Gamma_n \text{ is non-singular,} \quad \underline{a}_n = \Gamma_n^{-1} \underline{\gamma}_n. \tag{2.7}$$

It remains to show the predictor $\widehat{Y}$ is uniquely determined also when Γ_n is singular. Assume that Γ_n is singular and that $\underline{\gamma}_n = \Gamma_n \underline{a}_n^{(i)}$ for $i = 1, 2$. Let $\widehat{Y}^{(i)}$ for $i = 1, 2$ be the corresponding predictors. Then

$$\mathrm{Var}(\widehat{Y}^{(1)} - \widehat{Y}^{(2)}) = (\underline{a}^{(1)} - \underline{a}^{(2)})^t \Gamma_n (\underline{a}^{(1)} - \underline{a}^{(2)}) = (\underline{a}^{(1)} - \underline{a}^{(2)})^t (\underline{\gamma}_n - \underline{\gamma}_n) = 0,$$

from which it follows that $\widehat{Y}^{(1)} = \widehat{Y}^{(2)}$.

Note that if Γ_n is singular, then $\underline{a}^{(1)}$ and $\underline{a}^{(2)}$ may be different; all solutions $\underline{a}$ give the same predictor.

Furthermore, it follows from the above that the following result (known as the projection theorem) holds:

Theorem 2.5 (Projection Theorem)**.** *Let $\widehat{Y}$ be the linear predictor of Y based on $W_1, \ldots, W_n$, then $\widehat{Y}$ satisfies:*

$$Cov(\widehat{Y} - Y, W_i) = 0, \quad \textit{for} \quad i = 1, \ldots, n. \tag{2.8}$$

Proof Without loss of generality, let $\mu = \mu_1 = \ldots = \mu_n = 0$. From above, the linear predictor $\widehat{Y} = \sum_{j=1}^{n} a_j W_{n-j+1}$ satisfies $\frac{\partial S}{\partial a_i} = 0$, which is equivalent to

$$0 = \mathbb{E}\left[W_{n-i+1}(Y - \widehat{Y})\right] \qquad i = 1, \ldots, n$$

so that

$$0 = \mathbb{E}\left[W_i\left(Y - \widehat{Y}\right)\right] = \mathrm{Cov}(W_i, Y - \widehat{Y}) \qquad \forall i = 1, \ldots, n.$$

□

Since the problem with arbitrary known means is equivalent, from now on, only $\mu = \mu_1 = \ldots = \mu_n = 0$ will be considered. Consider the *mean-square prediction error*

$$v_n := \mathbb{E}\left[(\widehat{Y} - Y)^2\right] = \mathrm{Var}(\widehat{Y} - Y).$$

If Γ_n is non-singular, then it follows from equation ((2.8)) that

$$\mathrm{Var}(Y) = \mathrm{Var}(Y - \widehat{Y} + \widehat{Y}) = \mathrm{Var}(Y - \widehat{Y}) + \mathrm{Var}(\widehat{Y}).$$

It follows that

$$v_n = \mathrm{Var}(Y - \widehat{Y}) = \mathrm{Var}(Y) - \mathrm{Var}(\hat{Y}) = \mathrm{Var}(Y) - \underline{a}_n^t \Gamma_n \underline{a}_n = \mathrm{Var}(Y) - \underline{\gamma}_n^t \Gamma_n^{-1} \underline{\gamma}_n. \tag{2.9}$$

Now consider all random variables mean zero, predicting Y from $W_1, \ldots, W_n$. The linear predictors are of the form

$$\widehat{Y} = a_1 W_n + \ldots + a_n W_1, \tag{2.10}$$

Let

$$\mathcal{M} = \overline{\{a_1 W_n + \ldots + a_n W_1, \quad \underline{a} = (a_1, \ldots, a_n)^t \in \mathbb{R}^n\}}$$

and

$$\mathcal{H} = \overline{\{bY + a_1 W_n + \ldots + a_n W_1, \quad b \in \mathbb{R}, \quad \underline{a} = (a_1, \ldots, a_n)^t \in \mathbb{R}^n\}}$$

where the notation $\overline{\mathcal{A}}$ is used to denote the closure of the space $\mathcal{A}$. The predictor $\widehat{Y}$ may be considered as the point in $\mathcal{M}$ closest to Y, where distances in $\mathcal{H}$ are measured in terms of variances. A Hilbert space structure may be introduced to $\mathcal{H}$ by taking the inner product to be the covariance;

$$\langle X, Y \rangle = \mathrm{Cov}(X, Y).$$

It follows that for X and Z two random variables in $\mathcal{H}$,

$$X \perp Z \Leftrightarrow \mathrm{Cov}(X, Y) = 0; \qquad X \perp Z \Leftrightarrow \mathrm{Var}(X + Z) = \mathrm{Var}(X) + \mathrm{Var}(Z).$$

In this framework, Equation (2.8) states that $\hat{Y}$ is determined by $Y - \hat{Y} \perp \mathcal{M}$, hence Theorem 2.5 is referred to as the projection theorem.

It is clear that $\langle X, Y \rangle := \mathrm{Cov}(X, Y)$ satisfies the hypotheses for an inner product and that the spaces $\mathcal{H}$ and $\mathcal{M}$ are Hilbert spaces. The following notation will be used throughout: for a collection of random variables $\mathcal{C}$, $\mathcal{M}(\mathcal{C})$, or more simply $\mathcal{M}$ when the collection $\mathcal{C}$ is understood from the context, will denote the Hilbert space spanned by the variables $\mathcal{C}$. If the collection of variables is $\mathcal{F}_{a:b}^{(X)}$, then the Hilbert space will be denoted by $\mathcal{M}_{a:b}^{(X)}$. The notation $P_{\mathcal{M}(\mathcal{C})}(Y)$ or $P_{\mathcal{M}} Y$ when $\mathcal{C}$ is understood from the context will be used interchangeably to denote exactly the same thing; the projection of the random variable Y onto the Hilbert space $\mathcal{M}(\mathcal{C})$. This is also equal to $P_{\mathcal{C}}(Y)$ defined earlier. Note that:

$$\widehat{Y} = P_{\mathcal{F}_{1:n}^{(W)}} Y = P_{\mathcal{M}_{1:n}^{(W)}} Y.$$

The norm for the Hilbert space $\| \cdot \|$ is defined by: $\|X\| = \sqrt{\mathrm{Var}(X)}$. In particular,

$$\|Y - \widehat{Y}\|^2 = \mathrm{Var}(Y - \widehat{Y}).$$

The Hilbert space $\mathcal{M}_{1:n}^{(W)}$ is called the *Hilbert space spanned by* $W_1, \ldots, W_n$.

2.3 Partial correlation

Let Y_1 and Y_2 be two random variables. The strength of the *linear* relationship between them may be measured by the correlation coefficient

$$\rho(Y_1, Y_2) := \frac{\mathrm{Cov}(Y_1, Y_2)}{\sqrt{\mathrm{Var}(Y_1)\mathrm{Var}(Y_2)}}.$$

In many situations, a large correlation between Y_1 and Y_2 can be explained through other variables, $W_1, \ldots, W_k$.

Suppose that Y_1 and Y_2 are random variables (for example, two time points in a time series) and $W_1, \ldots, W_k$ are other random variables (indicating, for example, other time points in the series). Let

$$\widehat{Y}_1 = P_{\overline{\mathrm{spa}}\{1, W_1, \ldots, W_k\}}(Y_1) \qquad \text{and} \qquad \widehat{Y}_2 = P_{\overline{\mathrm{spa}}\{1, W_1, \ldots, W_k\}}(Y_2)$$

where spa denotes 'span' and the overline denotes closure. Therefore $\widehat{Y}_1$ is simply the projection of Y_1 onto $(1, W_1, \ldots, W_k)$, and similiarly for $\widehat{Y}_2$.

Definition 2.6 (Multiple Correlation Coefficient, Partial Correlation Coefficient)**.** *Let Y_1 and $W_1, \ldots, W_k$ be random variables. The* multiple correlation coefficient *between Y_1 and $W_1, \ldots, W_k$ is defined by $\rho(Y_1, \widehat{Y}_1)$. Let Y_1, Y_2 and $W_1, \ldots, W_k$ be random variables. The* partial correlation coefficient *of Y_1 and Y_2 with respect to $W_1, \ldots, W_k$ is defined by*

$$\alpha(Y_1, Y_2) := \rho(Y_1 - \widehat{Y}_1, Y_2 - \widehat{Y}_2). \tag{2.11}$$

Example 2.1 (Partial correlation with respect to a single variable)**.**

When $k = 1$ (the partial correlation with respect to one variable W), then $\widehat{Y}_1 = a_1 + b_1 W$ and $\widehat{Y}_2 = a_2 + b_2 W$. Furthermore, from the projection theorem, $\mathrm{Cov}(Y_1, W) = \mathrm{Cov}(\widehat{Y}_1, W) = b_1 \mathrm{Var}(W)$ and $\mathrm{Cov}(Y_2, W) = b_2 \mathrm{Var}(W)$, so $\rho(Y_1, W) = \frac{b_1\sqrt{\mathrm{Var}(W)}}{\sqrt{\mathrm{Var}(Y_1)}}$ and $\rho(Y_2, W) = \frac{b_2\sqrt{\mathrm{Var}(W)}}{\sqrt{\mathrm{Var}(Y_2)}}$. Formula (2.11) reduces to:

$$\begin{aligned}
\alpha(Y_1, Y_2) &= \frac{\mathrm{Cov}(Y_1, Y_2) + \mathrm{Cov}(\widehat{Y}_1, \widehat{Y}_2) - \mathrm{Cov}(Y_1, \widehat{Y}_2) - \mathrm{Cov}(Y_2, \widehat{Y}_1)}{\sqrt{\mathrm{Var}(Y_1) + \mathrm{Var}(\widehat{Y}_1) - 2\mathrm{Cov}(Y_1, \widehat{Y}_1)}\sqrt{\mathrm{Var}(Y_2) + \mathrm{Var}(\widehat{Y}_2) - 2\mathrm{Cov}(Y_2, \widehat{Y}_2)}} \\
&= \frac{\mathrm{Cov}(Y_1, Y_2) - b_1 b_2 \mathrm{Var}(W)}{\sqrt{\mathrm{Var}(Y_1) - b_1^2 \mathrm{Var}(W)}\sqrt{\mathrm{Var}(Y_2) - b_2^2 \mathrm{Var}(W)}} \\
&= \frac{\rho(Y_1, Y_2) - \rho(Y_1, W)\rho(Y_2, W)}{\sqrt{1 - \rho(Y_1, W)^2)(1 - \rho(Y_2, W)^2)}}
\end{aligned}$$

provided $|\rho(Y_1, W)| < 1$ and $|\rho(Y_2, W)| < 1$. □

Example 2.2. Let

$$Y_1 = W + \widetilde{W}_1 \quad \textit{and} \quad Y_2 = W + \widetilde{W}_2,$$

where $W, \widetilde{W}_1, \widetilde{W}_2$ are independent random variables, each with mean 0. Then

$$\widehat{Y}_k = P_{\overline{spa}\{W\}}(Y_k) = P_{\overline{spa}\{W\}}(W + \widetilde{W}_k) = P_{\overline{spa}\{W\}}(W) + P_{\overline{spa}\{W\}}\widetilde{(W}_k) = W + 0 = W,$$

from which

$$\alpha(Y_1, Y_2) = \rho(\widetilde{W}_1, \widetilde{W}_2) = 0.$$

□

Example 2.3. This example shows that uncorrelated variables may, in fact, be completely partially correlated. Let Y_1 and Y_2 be independent, each with mean 0 and the same variance σ^2. Let $W = Y_1 + Y_2$. Then

$$\widehat{Y}_1 = \widehat{Y}_2 = \frac{W}{2} = \frac{Y_1 + Y_2}{2},$$

from which

$$Y_1 - \hat{Y}_1 = \frac{Y_1 - Y_2}{2} = -(Y_2 - \hat{Y}_2).$$

It follows that $\alpha(Y_1, Y_2) = -1$.

□

2.4 Prediction for Stationary Time Series

Let $\{X_t\}$ be a stationary time series with mean 0 and autocovariance function$\gamma(.)$ and consider the problem of predicting $Y = X_{n+1}$ based on $X_1, \ldots, X_n$. Then, denoting the predictor by $\widehat{X}_{n+1}$, define $\phi_{n,j}$ such that

$$\widehat{X}_{n+1} = \sum_{j=1}^{n} \phi_{n,j} X_{n+1-j}.$$

In the notation of before,

$$\Gamma_{n;ij} = \mathrm{Cov}(X_{n+1-i}, X_{n+1-j}) = \gamma(|i-j|)$$

and

$$\gamma_j = \mathrm{Cov}(X_{n+1}, X_{n+1-j}) = \gamma(j).$$

For $\gamma(0) > 0$ and $\lim_{h \to +\infty} \gamma(h) = 0$, Γ_n is non-singular (exercise).

Using notation from before, it is clear that:

$$P_{\overline{\mathrm{spa}}\{X_2, \ldots, X_h\}}(X_{h+1}) = \sum_{i=1}^{h-1} \phi_{h-1,i} X_{h+1-i}.$$

Notation The following notation will be used: $P_h(X_{t+h})$ will be used to denote $P_{\mathcal{F}_{1:t}^{(X)}}(X_{t+h})$, the notation $\widehat{X}_{t+1}$ will be used for the one step predictor $P_1(X_{t+1}) = P_{\mathcal{F}_{1:t}^{(X)}}(X_{t+1})$.

2.5 Partial autocorrelation

Definition 2.7. *Let $\{X_t, t \in \mathbb{Z}\}$ be a zero-mean stationary time series. The* partial autocorrelation function (PACF) *of $\{X_t\}$ is defined by*

$$\begin{cases} \alpha(0) = 1 \\ \alpha(1) = \rho(1) \\ \alpha(h) = \rho(X_{h+1} - P_{\overline{spa}\{X_2,\ldots,X_h\}}(X_{h+1}), X_1 - P_{\overline{spa}\{X_2,\ldots,X_h\}}(X_1)) \qquad h \geq 2. \end{cases}$$

Example 2.4 (AR(p) Process)**.**

An AR(p) process is defined by:

$$X_t - \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_t \qquad \{\epsilon_t\} \sim WN(0, \sigma^2).$$

In the next section, we discuss conditions under which this process is well defined. If this equation defines a stationary process with $\mathbb{E}[X_t] = 0$ and $\mathrm{Var}(X_t) < +\infty$ such that $\mathrm{Corr}(X_s, \epsilon_t) = 0$ for $t \geq s+1$, then it follows almost directly from the definition that the partial autocorrelation function $\alpha(h)$ for an AR(p) process is equal to 0 for $|h| > p$.

2.6 The Wold decomposition

Let $\{X_t : t \in \mathbb{Z}\}$ be a zero-mean stationary time series. Let

$$\mathcal{M}_n = \overline{\mathrm{spa}}\{X_t,\ t \leq n\}, \quad \mathcal{M}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{M}_n, \qquad \sigma^2 = \mathbb{E}\left[|X_{n+1} - P_{\mathcal{M}_n}(X_{n+1})|^2\right]$$

Definition 2.8. *The process $\{X_t\}$ is called* deterministic *if $\sigma^2 = 0$, or equivalently if $X_t \in \mathcal{M}_{-\infty}$ $\forall t \in \mathbb{Z}$. The process $\{X_t\}$ is called* purely non-deterministic *if*

$$\mathcal{M}_{-\infty} = \{0\}.$$

Theorem 2.9 (The Wold decomposition)**.** *Let $\{X_t : t \in \mathbb{Z}\}$ be a zero-mean stationary time series, with $\mathcal{M}_n$, $\mathcal{M}_{-\infty}$ and σ^2 defined above. Suppose $\sigma^2 > 0$. Then X_t can be expressed as*

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + V_t, \tag{2.12}$$

where

1. $\psi_0 = 1 \quad$ *and* $\quad \sum_{j=0}^{\infty} \psi_j^2 < \infty$;

2. $\{\epsilon_t\} \sim WN(0, \sigma^2)$;

3. $\epsilon_t \in \mathcal{M}_t \quad$ *for each* $t \in \mathbb{Z}$;

4. $\mathbb{E}[\epsilon_t V_s] = 0 \quad$ *for all* $s, t \in \mathbb{Z}$;

5. $V_t \in \mathcal{M}_{-\infty} \quad$ *for each* $t \in \mathbb{Z}$;

6. $\{V_t\}$ *is* deterministic, *where the definition of 'deterministic' is given in Definition 2.8.*

The sequences $\{\psi_j\}$, $\{\epsilon_t\}$ *and* $\{V_t\}$ *are uniquely determined by* $\{X_t\}$ *and the conditions.*

Note The final two statements are not the same, since $\mathcal{M}_{-\infty}$ is defined in terms of $\{X_t\}$, not $\{V_t\}$.

Proof of Theorem 2.9 Firstly, it is shown that the sequences defined by:

$$\begin{cases} \epsilon_t = X_t - P_{\mathcal{M}_{t-1}}(X_t) \\ \psi_j = \frac{1}{\sigma^2}\langle X_t, \epsilon_{t-j}\rangle \\ V_t = X_t - \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \end{cases} \tag{2.13}$$

satisfy Equation (2.12) along with the six conditions of the theorem. The proof is then completed by establishing the uniqueness of the thre sequences.

Firstly, ϵ_t defined in Equation (2.13) is an element of $\mathcal{M}_t$ and is orthogonal to $\mathcal{M}_{t-1}$, by definition. It follows that

$$\epsilon_t \in \mathcal{M}_{t-1}^{\perp} \subset \mathcal{M}_{t-2}^{\perp} \subset \dots$$

so that, for all $s < t$, $\mathbb{E}[\epsilon_s \epsilon_t] = 0$. The second and third conditions of the theorem are therefore established. Next,

$$P_{\overline{\text{spa}}\{\epsilon_j : j \le t\}}(X_t) = \psi_j \epsilon_{t-j} \tag{2.14}$$

where ψ_j is defined by Equation (2.13) and $\sum_{j=1}^{\infty} \psi_j^2 < +\infty$. The coefficients ψ_j are independent of t by stationarity and

$$\psi_0 = \frac{1}{\sigma^2}\langle X_t, X_t - P_{\mathcal{M}_{t-1}}(X_t)\rangle = \frac{1}{\sigma^2}\|X_t - P_{\mathcal{M}_{t-1}}(X_t)\|^2 = 1.$$

The third equation of (2.13) together with Equation (2.14) give:

$$\langle V_t, \epsilon_s\rangle = 0 \qquad \forall s \le t.$$

For $s > t$, $\epsilon_s \in \mathcal{M}_{s-1}^{\perp} \subset \mathcal{M}_t^{\perp}$. Since $V_t \in \mathcal{M}_t$, it follows that

$$\langle V_t, \epsilon_s\rangle = 0 \qquad \forall s > t.$$

To establish the last two conditions, it is sufficient to show that

$$\overline{\text{spa}}\{V_j : j \leq t\} = \mathcal{M}_{-\infty} \qquad \forall t \in \mathbb{Z}. \tag{2.15}$$

Since $V_t \in \mathcal{M}_t = \mathcal{M}_{t-1} \oplus \overline{\text{spa}}\{\epsilon_t\}$ and since $\langle V_t, \epsilon_t \rangle = 0$, it follows that $V_t \in \mathcal{M}_{t-1}$. Using inductively $\mathcal{M}_s = \mathcal{M}_{s-1} \oplus \overline{\text{spa}}\{\epsilon_s\}$, it follows that $V_t \in \mathcal{M}_{t-j}$ for all $j \geq 0$ and hence $V_t \in \mathcal{M}_{-\infty}$. It follows that

$$\overline{\text{spa}}\{V_j : j \leq t\} \subseteq \mathcal{M}_{-\infty} \qquad \forall t \in \mathbb{Z}.$$

It follows from the third equation of (2.13) that:

$$\mathcal{M}_t = \overline{\text{spa}}\{\epsilon_j : j \leq t\} \oplus \overline{\text{spa}}\{V_j : j \leq t\}.$$

If $Y \in \mathcal{M}_{-\infty}$, then $Y \in \mathcal{M}_{s-1}$ for each $s \in \mathbb{Z}$ and therefore $\langle Y, \epsilon_s \rangle = 0$ for each $s \in \mathbb{Z}$. It follows that $Y \in \overline{\text{spa}}\{V_j : j \leq t\}$. This implies that

$$\mathcal{M}_{-\infty} \subseteq \overline{\text{spa}}\{V_j : j \leq t\} \qquad \forall t \in \mathbb{Z}.$$

This establishes Equation (2.15) and hence the existence part of the result.

To establish uniqueness: let $\{\epsilon_t\}$ and $\{V_t\}$ be any two sequences with the desired properties. Then

$$\mathcal{M}_{t-1} \subset \overline{\text{spa}}\{\epsilon_j : j \leq t-1\} \oplus \overline{\text{spa}}\{V_j : j \leq t-1\}$$

from which it follows that $\epsilon_t \perp \mathcal{M}_{t-1}$ (using the second and fourth property). From projecting each side of Equation (2.12) onto $\mathcal{M}_{t-1}$ and subtracting the resulting equation from (2.12), it follows that $\{\epsilon_t\}$ satisfies the first equation of (2.13). By taking inner products of each side of equation (2.12) with ϵ_{t-j}, it follows that ψ_j must satisfy the second of (2.13). Finally, for (2.12) to hold, it is clear that V satisfies the third of (2.13), hence uniqueness has been established. □