

2.5 Forecasting for Stationary Processes

Forecasting for stationary processes utilizes the autocovariance structure of the process to optimize predictions. Suppose we wish to forecast X_{n+h} given the observed data $x_{1:n} = (x_1, \dots, x_n)$. We focus on *linear* predictors, that is, linear combinations of past values.

We will focus on model-free procedures, and utilize moment properties rather than distributional properties of $\{X_t\}$.

2.5 Forecasting for Stationary Processes

We will again utilize the minimum Mean-Square Error (MSE) criterion, and choose constants $a_{0:n} = (a_0, a_1, \dots, a_n)^\top$ to minimize

$$\mathbb{E} \left[\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i X_{n-i+1} \right)^2 \middle| X_{1:n} = x_{1:n} \right]$$

We denote the linear predictor using the n -data prediction operator notation

$$a_0 + \sum_{i=1}^n a_i X_{n-i+1} = P_n X_{n+h}.$$

2.5 Forecasting for Stationary Processes

By differentiation, we need to solve simultaneously

$$\mathbb{E} \left[\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i x_{n-i+1} \right) \right] = 0$$

and, for $j = 1, \dots, n$

$$\mathbb{E} \left[\left(X_{n+h} - a_0 - \sum_{i=1}^n a_i x_{n-i+1} \right) x_{n-j+1} \right] = 0$$

From the first equation, we see that we must choose

$$a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right)$$

where $\mu = \mathbb{E}[X_t]$.

2.5 Forecasting for Stationary Processes

In vector notation, the remaining system of n simultaneously equations is written

$$\Gamma_n a = \gamma_{h:(h+n-1)} \quad (1)$$

where

$$\Gamma_n = [\gamma_X(i-j)]_{ij} \quad (n \times n)$$

$$a = (a_0, a_1, \dots, a_n)^\top \quad (n \times 1)$$

$$\gamma_{h:(h+n-1)} = (\gamma_X(h), \dots, \gamma_X(h+n-1))^\top \quad (n \times 1)$$

with $a_0 \equiv a_0(a_1, \dots, a_n)$ given by the previous expression. This is a **linear system** in $(a_1, \dots, a_n)$; in general the solution will also depend on h .

2.5 Forecasting for Stationary Processes

Denote the solution to (1) by

$$a(n, h) = (a(n, 1), \dots, a(n, n))^T$$

so that

$$P_n X_{n+h} = \mu + \sum_{i=1}^n a(n, i)(X_{n-i+1} - \mu).$$

Note that, by construction

$$\mathbb{E}[(X_{n+h} - P_n X_{n+h})] = 0$$

$$\mathbb{E}[(X_{n+h} - P_n X_{n+h})X_j] = 0 \quad j = 1, \dots, n$$

The minimum value of the MSE achieved by $a(n, h)$ is

$$\mathbb{E}[(X_{n+h} - P_n X_{n+h})^2] = \gamma_X(0) - a(n, h)^T \gamma_{h:(h+n-1)}$$

2.5 Forecasting for Stationary Processes

Recall that the formula

$$\Gamma_n a = \gamma_{h:(h+n-1)}$$

depends only on the autocovariance function. Direct solution of this equation is possible for most practical situations

$$a(n, h) = \Gamma_n^{-1} \gamma_{h:(h+n-1)}$$

can be computed using `solve` in R. However, the form of Γ_n (symmetric, Toeplitz) can make it relatively easy to invert.

2.5 Forecasting for Stationary Processes

Example (AR(1))

Recall that

$$[\Gamma_n]_{ij} = \phi^{|i-j|}$$

so that

$$\gamma_{h:(h+n-1)} = (\phi^h, \phi^{h+1}, \dots, \phi^{h+n-1})^\top$$

Then the matrix inversion can be carried out analytically, yielding

$$a(n, h) = (\phi^h, 0, 0, \dots, 0)^\top$$

which is an intuitively reasonable result. It follows that

$$P_n X_{n+1} = \phi X_n \qquad P_n X_{n+2} = \phi^2 X_n$$

and so on.

2.5 Forecasting for Stationary Processes

Recall a general linear prediction framework: let $Y, W_1, \dots, W_n$ be random variables with $\mathbb{E}[Y] = \mu_Y, \mathbb{E}[W_i] = \mu_i$, $\mathbb{V}ar[Y], \mathbb{V}ar[W_i], \mathbb{C}ov[Y, W_i], \mathbb{C}ov[W_i, W_j] < \infty$. Write $W_{n:1} = (W_1, \dots, W_n)^\top$, and

$$\begin{aligned}\gamma &= \mathbb{C}ov(Y, W_{n:1}) && (n \times 1) \\ \Gamma &= \mathbb{C}ov(W_{n:1}, W_{n:1}) && (n \times n)\end{aligned}$$

Denote the *best linear predictor* of Y , given $W_1, \dots, W_n$ by

$$\hat{P}(Y|W) = \mu_Y + a^\top (W_{n:1} - \mu_{n:1})$$

where a solves $\Gamma a = \gamma$. The minimum MSE is then

$$\mathbb{E}[(Y - \hat{P}(Y|W))^2] = \mathbb{V}ar[Y] - a^\top \gamma.$$

2.5 Forecasting for Stationary Processes

Let U and V be random variables with $\mathbb{E}[U^2], \mathbb{E}[V^2] < \infty$, and $\beta, \alpha_1, \dots, \alpha_n$ be constants. Then

1. $\hat{P}(U|W_{n:1}) = \mathbb{E}[U] + a^\top(W_{n:1} - \mu_{n:1})$.
2. $\mathbb{E}[U - \hat{P}(U|W_{n:1})] = \mathbb{E}[(U - \hat{P}(U|W_{n:1}))W_{n:1}] = \mathbf{0}$.
3. $\mathbb{E}[(U - \hat{P}(U|W_{n:1}))^2] = \mathbb{V}ar[U] - a^\top \gamma$.
4. $\hat{P}$ is a linear operator

$$\hat{P}(\alpha_1 U + \alpha_2 V + \beta | W_{n:1}) = \alpha_1 \hat{P}(U|W_{n:1}) + \alpha_2 \hat{P}(V|W_{n:1}) + \beta.$$

5. $\hat{P}(\alpha^\top W_{n:1} + \beta | W_{n:1}) = \alpha^\top W_{n:1} + \beta$.
6. If $\gamma = 0$, $\hat{P}(U|W_{n:1}) = \mathbb{E}[U]$.
7. Coherent iterated prediction:
$$\hat{P}(U|W) = \hat{P}(\hat{P}(U|W_{n:1}, W_{n:1}^*)).$$

2.5 Forecasting for Stationary Processes

Example (AR(p) process)

If $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and

$$\Phi(B)X_t = Z_t$$

such that $\{X_t\}$ is stationary and causal. Then for each n

$$\begin{aligned} P_n X_{n+1} &= P_n [\Phi(B)X_n + Z_n] = P_n [\phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} + Z_n] \\ &= \phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} \end{aligned}$$

by 4., 5. and 6.

2.5 Forecasting for Stationary Processes

Levinson-Durbin Algorithm

For a stationary process $\{X_t\}$, suppose we write

$$P_n X_{n+1} = \sum_{j=1}^n \phi_{n,j} X_{n-j+1} = \underline{\phi}_n^\top X_{1:n}$$

say as the general form of a linear predictor. By the arguments above, the optimal MSE choice for $\underline{\phi}_n$ is

$$\underline{\phi}_n = \Gamma_n^{-1} \gamma_{1:n}$$

where

$$[\Gamma_n]_{ij} = \gamma_X(|i-j|) \quad \gamma_{1:n} = (\gamma_X(1), \dots, \gamma_X(n))^\top.$$

Let the minimum MSE be denoted

$$v_n = \mathbb{E}[(X_{n+1} - P_n X_{n+1})^2] = \gamma_X(0) - \underline{\phi}_n^\top \gamma_{1:n}.$$

2.5 Forecasting for Stationary Processes

We seek a recursion for ϕ_n ; suppose ϕ_{n-1} and v_{n-1} have been computed, and define ϕ_n by setting the n th component as

$$\phi_{n,n} = \frac{1}{v_{n-1}} \left[\gamma_X(n) - \phi_{n-1}^\top \gamma_{(n-1):1} \right]$$

and the first $n - 1$ components via

$$\phi_{n,1:(n-1)} = \phi_{n-1} - \phi_{n,n} \phi_{n-1}^R$$

where

$$\phi_{n-1}^R = (\phi_{n-1,n-1}, \dots, \phi_{n-1,1})^\top$$

is the reversed version of ϕ_{n-1} . Finally, set

$$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

2.5 Forecasting for Stationary Processes

To initialize the recursion, set

$$\phi_{1,1} = \rho_X(1) = \frac{\gamma_X(1)}{\gamma_X(0)} \quad v_0 = \gamma_X(0) \quad v_1 = \gamma_X(0)(1 - \phi_{1,1}^2)$$

To see that the recursion produces a valid solution, first set R_n to be the autocorrelation matrix $R_n = \Gamma_n / \gamma_X(0)$, and define

$$\rho_{1:n} = \frac{1}{\gamma_X(0)} \gamma_{1:n} = (\rho_X(1), \dots, \rho_X(n))^T.$$

We verify the inductive step

$$R_n \phi_n = \rho_{1:n} \quad \Longrightarrow \quad R_{n+1} \phi_n = \rho_{1:(n+1)}$$

2.5 Forecasting for Stationary Processes

Clearly the relation holds for $n = 1$, as

$$R_1 = 1 \quad \phi_1 = \phi_{1,1} = \rho_X(1) \quad \rho_{1:1} = \rho_X(1).$$

We assume the relation holds for $n = k$, and verify that it holds for $n = k + 1$. Now

$$R_{k+1} = \begin{bmatrix} R_k & \rho_{k:1} \\ \{\rho_{k:1}\}^\top & 1 \end{bmatrix}$$

where

$$\rho_{k:1} = (\rho_X(k), \dots, \rho_X(1))^\top = \frac{1}{\gamma_X(0)} \gamma_{1:k}$$

2.5 Forecasting for Stationary Processes

Recall that

$$R_{k+1} = \frac{1}{\gamma_X(0)} \begin{bmatrix} \gamma_X(0) & \gamma_X(1) & \gamma_X(2) & \cdots & \gamma_X(k) \\ \gamma_X(1) & \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(k-1) \\ \gamma_X(2) & \gamma_X(1) & \gamma_X(0) & \cdots & \gamma_X(k-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma_X(k) & \gamma_X(k-1) & \gamma_X(k-2) & \cdots & \gamma_X(0) \end{bmatrix}$$

and note that

$$R_k \phi_k = \rho_{1:k} \quad \implies \quad R_k \phi_k^R = \rho_{k:1}.$$

2.5 Forecasting for Stationary Processes

$$\begin{aligned}
 R_{k+1}\phi_{k+1} &= \begin{bmatrix} R_k & \rho_{k:1} \\ \{\rho_{k:1}\}^\top & 1 \end{bmatrix} \begin{bmatrix} \phi_k - \phi_{k+1,k+1}\phi_k^R \\ \phi_{k+1,k+1} \end{bmatrix} \\
 &= \begin{bmatrix} R_k\phi_k - \phi_{k+1,k+1}R_k\phi_k^R + \phi_{k+1,k+1}\rho_{k:1} \\ \{\rho_{k:1}\}^\top \phi_k - \phi_{k+1,k+1} \{\rho_{k:1}\}^\top \phi_k^R + \phi_{k+1,k+1} \end{bmatrix} \\
 &= \begin{bmatrix} R_k\phi_k \\ \{\rho_{k:1}\}^\top \phi_k + \phi_{k+1,k+1}(1 - \{\rho_{k:1}\}^\top \phi_k^R) \end{bmatrix}
 \end{aligned}$$

as $R_k\phi_k^R = \rho_{k:1}$.

2.5 Forecasting for Stationary Processes

Now by definition

$$\phi_{k+1,k+1} = \frac{1}{v_k} \left[\gamma_X(k+1) - \phi_k^\top \gamma_{k:1} \right]$$

and by definition

$$\begin{aligned} v_k = \gamma_X(0) - \phi_k^\top \gamma_{1:k} &= \gamma_X(0) - \left\{ \phi_k^R \right\}^\top \gamma_{k:1} \\ &= \frac{1}{\gamma_X(0)} (1 - \{\rho_{k:1}\}^\top \phi_k^R) \end{aligned}$$

so therefore

$$\frac{1}{v_k} (1 - \{\rho_{k:1}\}^\top \phi_k^R) = \frac{1}{\gamma_X(0)}$$

2.5 Forecasting for Stationary Processes

Therefore

$$\phi_{k+1,k+1} = \frac{1}{\gamma_X(0)} \left[\gamma_X(k+1) - \phi_k^\top \gamma_{k:1} \right] = \rho_X(k+1) - \phi_k^\top \rho_{k:1}$$

and hence

$$\begin{aligned} R_{k+1} \phi_{k+1} &= \begin{bmatrix} R_k \phi_k \\ \{\rho_{k:1}\}^\top \phi_k + \rho_X(k+1) - \phi_k^\top \rho_{k:1} \end{bmatrix} \\ &= \begin{bmatrix} \rho_{1:k} \\ \rho_X(k+1) \end{bmatrix} \\ &= \rho_{1:(k+1)} \end{aligned}$$

and the recursion holds.

2.5 Forecasting for Stationary Processes

For the prediction MSE recursion: using the recursion for ϕ_n ,

$$\begin{aligned}v_n &= \mathbb{E}[(X_{n+1} - \phi_n^\top X_{n:1})^2] = \gamma_X(0) - \phi_n^\top \gamma_{1:n} \\&= \gamma_X(0) - \phi_{n-1}^\top \gamma_{1:(n-1)} + \phi_{n,n} \left\{ \phi_{n-1}^R \right\}^\top \gamma_{1:(n-1)} - \phi_{n,n} \gamma_X(n) \\&= v_{n-1} + \phi_{n,n} \left(\left\{ \phi_{n-1}^R \right\}^\top \gamma_{1:(n-1)} - \gamma_X(n) \right) \\&= v_{n-1} - \phi_{n,n}^2 (\gamma_X(0) - \phi_{n-1}^\top \gamma_{1:(n-1)}) \\&= v_{n-1} (1 - \phi_{n,n}^2).\end{aligned}$$

2.5 Forecasting for Stationary Processes

Partial Autocorrelation Function (PACF): Define the function $\alpha_X(h)$ by $\alpha_X(0) = 1$ and

$$\alpha_X(h) = \phi_{h,h} \quad h = 1, 2, \dots, n$$

where $\phi_{h,h}$ is the value returned on the h th step of the Levinson-Durbin recursion. This function is termed the *partial autocorrelation function*, or PACF, which is uniquely determined by the ACF γ_X .

Note also that as v_n is non-negative, and

$$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

we must have that $\phi_{n,n}^2 \leq 1$ for all n , and hence that

$$v_n \leq v_{n-1}$$

so that the optimal prediction MSEs for a **decreasing** sequence. 135

2.5 Forecasting for Stationary Processes

For random variables X, Y, Z , the *partial correlation* of X and Y , given Z is defined by

$$\mathbb{C}orr[(X - \mathbb{E}[X|Z]), (Y - E[Y|Z])]$$

that is, the correlation between the residuals from a regression of X on Z with those from a regression of Y on Z . For process $\{X_t\}$, if we let $X \equiv X_t$, $Y = X_{t+h}$, and $Z = X_{(t+1):(t+h-1)}$.

$$\begin{aligned}\alpha_X(h) &= \phi_{h,h} \\ &= \mathbb{C}orr[(X_t - \hat{P}[X_t|X_{(t+1):(t+h-1)}]), \\ &\quad (X_{t+h} - \hat{P}[X_{t+h}|X_{(t+1):(t+h-1)}])]\end{aligned}$$

The PACF computes the correlation between prediction residuals for X_t and X_{t+h} , using $X_{(t+1):(t+h-1)}$ to make the predictions.

2.5 Forecasting for Stationary Processes

Example (AR(p))

Suppose $\{X_t\} \sim \text{AR}(p)$, with parameters $\phi_1, \dots, \phi_p$, and consider $h > p$. Then

$$\hat{P}[X_{t+h}|X_{(t+1):(t+h-1)}] = \phi_1 X_{t+h-1} + \dots + \phi_p X_{t+h-p}$$

and

$$\hat{P}[X_t|X_{(t+1):(t+h-1)}] = X_t$$

so therefore

$$X_{t+h} - \hat{P}[X_{t+h}|X_{(t+1):(t+h-1)}] = Z_{t+h}$$

$$X_t - \hat{P}[X_t|X_{(t+1):(t+h-1)}] = Z_t$$

and as $\text{Corr}[Z_t, Z_{t+h}] = 0$, the PACF is **zero** for $h > p$.

2.5 Forecasting for Stationary Processes

The Levinson-Durbin algorithm provides a method for solving the linear system

$$R_n \phi_n = \rho_{1:n}$$

by recursion, for arbitrary n , and given ACF $\gamma_X(h)$ without the need for large matrix inversion. It exploits the symmetric Toeplitz nature of an autocovariance matrix, and is an order n^2 algorithm.

Other algorithms for inverting this type of matrix are typically higher order; the Cholesky decomposition

$$R_n = L_n L_n^T$$

with L_n lower triangular is order n^3 . It does not exploit the Toeplitz nature of R_n .

2.5 Forecasting for Stationary Processes

Innovations Algorithm

Another approach to forecasting proves feasible even for the non-stationary case. Suppose that $\{X_t\}$ is zero mean process with $\mathbb{E}[|X_t|^2]$, and $\mathbb{E}[X_i X_j] = \kappa(i, j)$. Denote

$$\hat{X}_n = \begin{cases} 0 & n = 1 \\ P_{n-1}X_n & n = 2, 3, \dots \end{cases}$$

and let

$$U_n = X_n - \hat{X}_n$$

denote the one step prediction error. Then

$$U_{1:n} = A_n X_{1:n}$$

where A_n contains the coefficients $a_{i,j}$ that appear in the optimal linear predictor.

2.5 Forecasting for Stationary Processes

$$A_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{1,1} & 1 & 0 & \cdots & 0 \\ a_{2,2} & a_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix}$$

This lower triangular matrix is non-singular and let

$$A_n^{-1} = C_n$$

where C_n is also lower triangular.

2.5 Forecasting for Stationary Processes

Let

$$C_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix}$$

Then, recalling that $X_{1:n} = C_n U_{1:n}$,

$$X_{1:n} = C_n U_{1:n} = C_n (X_{1:n} - \hat{X}_{1:n}).$$

Also

$$\hat{X}_{1:n} = X_{1:n} - U_{1:n} = C_n U_{1:n} - U_{1:n} = \Theta_n (X_{1:n} - \hat{X}_{1:n})$$

2.5 Forecasting for Stationary Processes

$$\Theta_n = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 0 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}$$

Therefore

$$\hat{X}_{n+1} = \begin{cases} 0 & n = 0 \\ \sum_{j=1}^n \theta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}) & n = 1, 2, \dots \end{cases}$$

2.5 Forecasting for Stationary Processes

Therefore the one-step predictions

$$\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n, \dots$$

can be computed as weighted sums of previous prediction errors

$$X_1 - \hat{X}_1, X_2 - \hat{X}_2, \dots, X_n - \hat{X}_n, \dots$$

provided the entries in the matrix Θ_n are known. Note that by the previous results, the sequence $\{D_n\}$ defined by

$$D_n = X_n - \hat{X}_n$$

is an **uncorrelated** sequence. Another recursive approach

avoids the need to carry out matrix inversion.

2.5 Forecasting for Stationary Processes

Initialize $v_0 = \kappa(1, 1) = \mathbb{E}[X_1^2]$, then for each n set

$$\theta_{n,n-k} = \frac{1}{v_k} \left[\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right] \quad 0 \leq k < n$$

and

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j$$

That is, we compute

$$\begin{array}{cccc} v_0 & & & \\ \theta_{1,1} & v_1 & & \\ \theta_{2,2} & \theta_{2,1} & v_2 & \\ \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & v_3 \end{array}$$

and so on.

2.5 Forecasting for Stationary Processes

For h -step ahead prediction

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+j-1,j} (X_{n+h-j} - \hat{X}_{n+j-j})$$

and

$$v_{n+h} = \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+j-1,j}^2 v_{n+h-j-1}$$

In this case, the innovations algorithm must be run **forward** $h - 1$ steps.

2.5 Forecasting for Stationary Processes

Note: It is possible to extend the idea of optimal linear prediction from the finite case to the *infinite*, that is, to find constants $\{\beta_j\}$ such that

$$\tilde{P}X_{n+h} = \sum_{j=1}^{\infty} \beta_j X_{n+1-j}.$$

The infinite system of moment constraints

$$\mathbb{E}[(X_{n+h} - \tilde{P}X_{n+h})X_{n+1-i}] = 0 \quad i = 1, 2, \dots$$

yield the infinite system of equations

$$\sum_{j=1}^{\infty} \gamma_X(i-j)\beta_j = \gamma_X(h+i-1) \quad i = 1, 2,$$

2.5 Forecasting for Stationary Processes

Deterministic Processes and The Wold Decomposition

Any process $\{X_t\}$ where

$$X_n - \tilde{P}_{n-1}X_n = 0$$

in mean square, that is

$$\mathbb{E}[(X_n - \tilde{P}_{n-1}X_n)^2] = 0$$

for all n is termed **deterministic**

2.5 Forecasting for Stationary Processes

Example (Simple Deterministic Process)

Suppose A, B are zero mean and uncorrelated variables with variance σ^2 , and let

$$X_t = A \cos(\omega t) + B \sin(\omega t)$$

for some $\omega \in (0, \pi)$. Then for each integer n ,

$$X_n = 2 \cos(\omega) X_{n-1} - X_{n-2} = \tilde{P}_{n-1} X_n$$

say, and hence

$$X_n - \tilde{P}_{n-1} X_n = 0$$

2.5 Forecasting for Stationary Processes

The *Wold Decomposition* gives a fundamental representation of stationary processes. If $\{X_t\}$ is stationary and non-deterministic, then $\{X_t\}$ has the representation

$$X_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j} + V_t \quad (\text{Wold})$$

where

- $\{Z_t\} \sim \text{WN}(0, \sigma^2)$
- $\{V_t\}$ is deterministic
- $\{Z_t\}$ and $\{V_t\}$ are uncorrelated
- $\psi_0 = 1, \sum_{j=1}^{\infty} \psi_j^2 < \infty.$

2.5 Forecasting for Stationary Processes

We also have that

- $Z_t = \tilde{P}_t Z_t$ for all t
- $V_t = \tilde{P}_s V_t$ for all s, t .

and that the representation in (Wold) is unique. Furthermore

- for each t , $Z_t = X_t - \tilde{P}_{t-1} X_t$;
- for each j , ψ_j satisfies

$$\psi_j = \frac{\mathbb{E}[X_t Z_{t-j}]}{\mathbb{E}[Z_t^2]};$$

- for each t , $V_t = X_t - \sum_{j=1}^{\infty} \psi_j Z_{t-j}$.