Chapter 8

Computing the Predictor

Recall that, for a univariate time series, the one-step predictor $\widehat{X}_n$ which minimises

$$v_n := \mathbb{E}[|X_n - \widehat{X}_n|^2]$$

is given by $\widehat{X}_n = \sum_{j=1}^{n} \phi_{n,j} X_{n-j}$ where the vector $\phi_{n,\cdot}$ satisfies:

$$\phi_{n,\cdot} = \Gamma_n^{-1} \gamma_n,$$

Γ_n denotes the correlation matrix of the predictors $\Gamma_{n;i,j} = \mathrm{Cov}(X_{n-i}, X_{n-j})$ while $\gamma_{n,i} = \mathrm{Cov}(X_n, X_{n-i})$.

The formula for the predictor can easily be generalised to a multivariate time series. If n is small, the matrix Γ may be inverted and the predictor computed explicitly with good results, but as n increases, the computation of Γ_n^{-1} may not be feasible. There are several recursive methods available. Recursive methods assume that $\widehat{X}_n$ has been computed, based on $X_1, X_2, \ldots, X_{n-1}$ and then, when X_n is obtained, the predictor $\widehat{X}_{n+1}$ is computed based on $X_1, X_2, \ldots, X_n$. The new information obtained is $\widehat{\epsilon}_n := X_n - \widehat{X}_n$.

Recall the definition of $\phi_{n,j}$;

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

We also use v_n to denote the mean squared error of the predictor:

$$v_n := \mathbb{E}\left[\left|X_{n+1} - \widehat{X}_{n+1}\right|^2\right].$$

The index n of v_n refers to the time point from which the one-step-ahead prediction is made.

8.1 The Durbin-Levinson Algorithm

Assume now that $\underline{\phi}_{n-1}$ and v_{n-1} have been computed, from which $\widehat{X}_n$ was obtained. The *Durbin - Levinson* algorithm is a method for obtaining $\widehat{X}_{n+1}$ once, in addition, X_n is known.

Theorem 8.1 (The Durbin-Levinson Algorithm). *Let $\{X_t\}$ be a zero-mean stationary time series with ACVF γ satisfying:*

$$\gamma(0) > 0, \qquad \gamma(h) \to 0 \quad as \quad h \to \infty.$$

Then

$$\phi_{1,1} = \frac{\gamma(1)}{\gamma(0)} \qquad , v_0 = \gamma(0),$$

$$\phi_{n,n} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1} \tag{8.1}$$

$$\begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{n,n} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix} \tag{8.2}$$

and

$$v_n = v_{n-1}[1 - \phi_{n,n}^2]. \tag{8.3}$$

Proof Consider the two orthogonal subspaces $\mathcal{K}_1 = \mathcal{M}_{2:n}^{(X)}$ and $\mathcal{K}_2 = \overline{\text{spa}}\{X_1 - P_{\mathcal{K}_1}X_1\}$. We use $\mathcal{M}_{m:n}^{(X)}$ to denote the closure of the space spanned by $X_m, \ldots, X_n$, so that $\mathcal{K}_1$ is the closure of the space spanned by $X_2, \ldots, X_n$, while the space $\mathcal{K}_1$ is the subspace of the space spanned by $X_1, \ldots, X_n$ which is orthogonal to $\mathcal{K}_2$.

Using this notation, the predictor $\widehat{X}_{n+1}$ may be expressed as:

$$\widehat{X}_{n+1} = P_{\mathcal{K}_1}(X_{n+1}) + P_{\mathcal{K}_2}(X_{n+1}) \tag{8.4}$$

It follows from the construction that

$$P_{\mathcal{K}_2}(X_{n+1}) = a(X_1 - P_{\mathcal{K}_1}(X_1)) \tag{8.5}$$

for some constant a, and from the projection theorem,

$$0 = \mathbf{C}(X_{n+1} - P_{\mathcal{K}_2}(X_{n+1}), P_{\mathcal{K}_2}(X_{n+1})) = a\langle X_{n+1}, X_1 - P_{\mathcal{K}_1}(X_1)\rangle - a^2\|X_1 - P_{\mathcal{K}_1}(X_1)\|^2, \tag{8.6}$$

from which

$$a = \frac{\langle X_{n+1}, X_1 - P_{\mathcal{K}_1}X_1\rangle}{\|X_1 - P_{\mathcal{K}_1}X_1\|^2}. \tag{8.7}$$

By stationarity, and using the fact that the 'reversed' process has the same ACVF,

$$P_{\mathcal{K}_1}(X_1) = \sum_{j=1}^{n-1} \phi_{n-1,j}X_{j+1}, \tag{8.8}$$

$$P_{\mathcal{K}_1}(X_{n+1}) = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{n+1-j}. \tag{8.9}$$

and

$$\|X_1 - P_{\mathcal{K}_1}(X_1)\|^2 = \|X_{n+1} - P_{\mathcal{K}_1}(X_{n+1})\|^2 = \|X_n - \widehat{X}_n\|^2 = v_{n-1} \tag{8.10}$$

From Equations (8.4), (8.5), (8.8) and (8.9)), it follows that

$$\widehat{X}_{n+1} = aX_1 + \sum_{j=1}^{n-1} \left(\phi_{n-1,j} - a\phi_{n-1,n-j}\right) X_{n+1-j}. \tag{8.11}$$

From Equations (8.7) and (8.8), it follows that

$$a = \left(\langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} \phi_{n-1,j} \langle X_{n+1}, X_{j+1} \rangle \right) v_{n-1}^{-1} = \left(\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right) v_{n-1}^{-1}.$$

Since $\underline{\phi}_n = \Gamma_n^{-1} \underline{\gamma}_n$ and the matrix Γ_n is non singular, the representation

$$\widehat{X}_{n+1} = \sum_{j=1}^{n} \phi_{nj} X_{n+1-j} \tag{8.12}$$

is unique. Comparing coefficients in Equations (8.11) and (8.12) therefore gives

$$\begin{cases} \phi_{nn} = a \\ \phi_{nj} = \phi_{n-1,j} - a\phi_{n-1,j-1}, \quad j = 1, \ldots, n-1. \end{cases} \tag{8.13}$$

Equations (8.1) and (8.2) have therefore been established. It only remains to establish Equation (8.3).

Using $X_{n+1} - P_{\mathcal{K}_1}(X_{n+1}) = X_{n+1} - \widehat{X}_{n+1} + P_{\mathcal{K}_2}(X_{n+1})$ and, by the projection theorem, $X_{n+1} - \widehat{X}_{n+1} \perp P_{\mathcal{K}_2}(X_{n+1})$, the mean squared error of the predictor is:

$$\begin{aligned} v_n &= \|X_{n+1} - \widehat{X}_{n+1}\|^2 \\ &= \|X_{n+1} - P_{\mathcal{K}_1}(X_{n+1}) - P_{\mathcal{K}_2}(X_{n+1})\|^2 \\ &= \|X_{n+1} - P_{\mathcal{K}_1}(X_{n+1})\|^2 + \|P_{\mathcal{K}_2}(X_{n+1})\|^2 - 2\langle X_{n+1} - P_{\mathcal{K}_1}(X_{n+1}), P_{\mathcal{K}_2}(X_{n+1}) \rangle \\ &= \|X_n - \widehat{X}_n\|^2 + a^2 \|X_1 - P_{\mathcal{K}_1}(X_1)\|^2 - 2a^2 \|X_1 - P_{\mathcal{K}_1}(X_1)\|^2 \\ &= v_{n-1}(1 - a^2) \end{aligned}$$

where Equation (8.12), the orthogonality of $\mathcal{K}_1$ and $\mathcal{K}_2$ and $P_{\mathcal{K}_2}(X_{n+1}) = a(X_1 - P_{\mathcal{K}_1}(X_1))$ and Equation (8.6) have been used. From Equation (8.7), it follows that

$$v_n = v_{n-1}(1 - a^2)$$

as required. □

Theorem 8.2. *Under the assumptions of Theorem 8.1, $\alpha(h) = \phi_{h,h}$ for $h \geq 1$.*

Proof Consider the proof of Theorem 8.1. From Equation (8.13) giving $a = \phi_{n,n}$ and Equation (8.7), it follows that

$$\phi_{n,n} = \frac{\langle X_{n+1}, X_1 - P_{\mathcal{K}_1}(X_1)\rangle}{\|X_1 - P_{\mathcal{K}_1}(X_1)\|^2}.$$

From the projection theorem, $X_1 - P_{\mathcal{K}_1}(X_1) \perp P_{\mathcal{K}_1}(X_{n+1})$, from which

$$\phi_{n,n} = \frac{\langle X_{n+1} - P_{\mathcal{K}_1}(X_{n+1}), X_1 - P_{\mathcal{K}_1}(X_1)\rangle}{\|X_1 - P_{\mathcal{K}_1}(X_1)\|^2} = \alpha(n).$$

□

Example 8.1 (AR(p) process)**.**

For an AR(p) process, it follows directly from the above that

$$\widehat{X}_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}, \quad \text{for} \quad n \geq p.$$

8.1.1 Multivariate Durbin-Levinson Algorithm

The projection theorem and predictors remain largely the same, replacing scalars with appropriate matrices. Let $\{\underline{X}_t : t \in \mathbb{Z}\}$ be an m-variate time series. Let $S_n := \overline{\text{spa}}(\underline{X}_1, \ldots, \underline{X}_n)$ and let $P(\underline{Y}|\mathcal{F}_n^{(X)})$ denote the projection of a random vector $\underline{Y}$ onto S_n. Then, by the projection theorem,

$$\begin{cases} P(\underline{Y}|\mathcal{F}_n^{(X)}) \in \mathcal{F}_n^{(X)}) \\ \underline{Y} - P(\underline{Y}|\mathcal{F}_n^{(X)}) \perp \underline{X}_{n+1-i} \quad i = 1, \ldots, n. \end{cases}$$

where

$$\underline{X} \perp \underline{Y} \Leftrightarrow \rho(X_i, Y_j) = 0 \qquad \forall i, j.$$

The best linear predictor $\widehat{\underline{X}}_{n+1}$ of $\underline{X}_{n+1}$ based on $\mathcal{F}_{1:n}^{(X)}$ is simply the vector of best linear predictors, each based on $(X_{t,j})_{t=1,\ldots,n;j=1,\ldots,m}$ and may be expressed as:

$$\widehat{\underline{X}}_{n+1} = \Phi_{n1}\underline{X}_n + \ldots + \Phi_{nn}\underline{X}_1$$

where $(\Phi_{jn})_{j=1}^n$ are $m \times m$ real valued matrices. The prediction equations are therefore:

$$\sum_{j=1}^{n} \Phi_{nj} K(n+1-j, n+1-i) = K(n+1, n+1-i) \qquad i = 1, \ldots, n \tag{8.14}$$

where

$$K(i,j) = \mathbb{E}[(\underline{X}_i - \mathbb{E}[\underline{X}_i])(\underline{X}_j - \mathbb{E}[\underline{X}_j])^t].$$

when $\underline{X}$ is stationary, $K(i,j) = \Gamma(i-j)$ and this reduces to

$$\sum_{j=1}^{n} \Phi_{nj} \Gamma(i-j) = \Gamma(i) \qquad i = 1, \ldots, n \tag{8.15}$$

The coefficients $\{\Phi_{nj}\}$ may be computed recursively from a multivariate version of the Durbin Levinson algorithm, which involves calculation of both the forward and backward predictors $P(\underline{X}_{n+1}|\mathcal{F}_{1:n}^{(X)})$ and $P(\underline{X}_0|\mathcal{F}_{1:n}^{(X)})$. Let $\Phi_{n1}, \ldots, \Phi_{nn}$ be the $m \times m$ coefficient matrices satisfying:

$$P(\underline{X}_0|\mathcal{F}_{1:n}^{(X)}) = \sum_{j=1}^{n} \widetilde{\Phi}_{nj}\underline{X}_j \qquad j = 1, 2, \ldots$$

Then

$$\sum_{j=1}^{n} \widetilde{\Phi}_{nj}\Gamma(j-i) = \Gamma(-i) \qquad i = 1, \ldots, n$$

Let V_n and $\widetilde{V}_n$ denote the two prediction error covariance matrices:

$$\begin{cases} V_n = \mathbb{E}[(\underline{X}_{n+1} - P(\underline{X}_{n+1}|\mathcal{F}_{1:n}^{(X)})(\underline{X}_{n+1} - P(\underline{X}_{n+1}|\mathcal{F}_{1:n}^{(X)})^t] \\ \widetilde{V}_n = \mathbb{E}[(\underline{X}_0 - P(\underline{X}_0|\mathcal{F}_{1:n}^{(X)})(\underline{X}_0 - P(\underline{X}_0|\mathcal{F}_{1:n}^{(X)})^t] \end{cases}$$

These satisfy:

$$\begin{cases} V_n = \mathbb{E}[(\underline{X}_{n+1} - P(\underline{X}_{n+1}|\mathcal{F}_{1:n}^{(X)})\underline{X}_{n+1}^t] = \Gamma(0) - \Phi_{n1}\Gamma(-1) - \ldots - \Phi_{nn}\Gamma(-n) \\ \widetilde{V}_n = \Gamma(0) - \widetilde{\Phi}_{n1}\Gamma(1) - \ldots - \widetilde{\Phi}_{nn}\Gamma(n) \end{cases} \tag{8.16}$$

Let

$$\begin{cases} \Delta_n := \mathbb{E}[(\underline{X}_{n+1} - P(\underline{X}_{n+1}|\mathcal{F}_{1:n}^{(X)}))\underline{X}_0^t] = \Gamma(n+1) - \Phi_{n1}\Gamma(n) - \ldots - \Phi_{nn}\Gamma(1) \\ \widetilde{\Delta}_n := \mathbb{E}[(\underline{X}_0 - P(\underline{X}_0|\mathcal{F}_{1:n}^{(X)}))\underline{X}_{n+1}^t] = \Gamma(-n-1) - \widetilde{\Phi}_{n1}\Gamma(-n) - \ldots - \widetilde{\Phi}_{nn}\Gamma(-1) \end{cases} \tag{8.17}$$

Proposition 8.3 (The Multivariate Durbin Levinson Algorithm)**.** *Let $\{\underline{X}_t : t \in \mathbb{Z}\}$ be a stationary m-variate linear time series, mean zero, with autocovariance function $\Gamma(h)$, such that covariance matrix of $vec(\underline{X}_1, \ldots, \underline{X}_n)$ is non-singular for each $n \geq 1$. Then the coefficients $\{\Phi_{nj}\}_{n\geq 1, j=1,\ldots,n}$ are given by:*

$$\begin{cases} \Phi_{nn} = \Delta_{n-1}\widetilde{V}_{n-1}^{-1} \\ \widetilde{\Phi}_{nn} = \widetilde{\Delta}_{n-1}V_{n-1}^{-1} \\ \Phi_{nk} = \Phi_{n-1,k} - \Phi_{nn}\widetilde{\Phi}_{n-1,n-k} \quad k = 1, \ldots, n-1 \\ \widetilde{\Phi}_{nk} = \widetilde{\Phi}_{n-1,k} - \widetilde{\Phi}_{nn}\Phi_{n-1,n-k} \quad k = 1, \ldots, n-1 \end{cases}$$

where V_n and $\widetilde{V}_n$ are defined by Equation (8.16) *and Δ_n and $\widetilde{\Delta}_n$ are defined by Equation* (8.17) *with*

$$\begin{cases} V_0 = \widetilde{V}_0 = \Gamma(0) \\ \Delta_0 = \widetilde{\Delta}_0^t = \Gamma(1). \end{cases}$$

Proof The proof closely parallels the univariate case. For $n = 1$, the result follows directly from Equation (8.15). For $n \geq 2$,

$$\underline{\widehat{X}}_{n+1} = P(X_{n+1}|\mathcal{F}_{2:n}^{(X)}) + A\underline{U}$$

where $\underline{U} = \underline{X}_1 - P(\underline{X}_1|\mathcal{F}_{2:n}^{(X)})$ and A is an $m \times m$ matrix that satisfies:

$$\underline{X}_{n+1} - A\underline{U} \perp \underline{U};$$

in other words,

$$\mathbb{E}\left[\underline{X}_{n+1}\underline{U}^t\right] = A\mathbb{E}\left[\underline{U}\underline{U}^t\right].$$

It follows from stationarity that

$$\begin{cases} P(\underline{X}_{n+1}|\mathcal{F}_{2:n}^{(X)}) = \Phi_{n-1,1}\underline{X}_1 + \ldots + \Phi_{n-1,n-1}\underline{X}_2 \\ \underline{U} = \underline{X}_1 - \Phi_{n-1,1}\underline{X}_2 - \ldots - \Phi_{n-1,n-1}\underline{X}_n \end{cases}$$

and

$$\mathbb{E}\left[\underline{U}\underline{U}^t\right] = \widetilde{V}_{n-1}.$$

It now follows from the foregoing that:

$$\begin{aligned} A &= \mathbb{E}\left[\underline{X}_n\underline{U}^t\right]\widetilde{V}_{n-1}^{-1} \\ &= \mathbb{E}\left[\left(\underline{X}_{n+1} - P(\underline{X}_{n+1}|\mathcal{F}_{2:n})\right)\underline{U}^t\right]\widetilde{V}_{n-1}^{-1} \\ &= \mathbb{E}\left[\left(\underline{X}_{n+1} - P(\underline{X}_{n+1}|\mathcal{F}_{2:n}^{(X)})\right)\underline{X}_1^t\right]\widetilde{V}_{n-1}^{-1} \\ &= \left(\Gamma(n) - \Phi_{n-1,1}\Gamma(n-1) - \ldots - \Phi_{n-1,n-1}\Gamma(1)\right)\widetilde{V}_{n-1}^{-1} \\ &= \Delta_{n-1}\widetilde{V}_{n-1}^{-1} \end{aligned}$$

from which it follows that

$$\underline{\widehat{X}}_{n+1} = A\underline{X}_1 + \sum_{j=1}^{n-1}\left(\Phi_{n-1,j} - A\Phi_{n-1,n-j}\right)\underline{X}_{n+1-j},$$

which proves half of the recursions. The other half is established by a symmetric argument. □

8.2 The Innovations Algorithm

It is equivalent, and natural, to consider predictors which are linear combinations of the *innovations* $X_1 - \widehat{X}_1, \ldots, X_n - \widehat{X}_n$. This is equivalent since

$$\overline{\text{spa}}\{X_1, X_2, \ldots, X_n\} = \overline{\text{spa}}\{X_1 - \widehat{X}_1, \ldots, X_n - \widehat{X}_n\}.$$

The *innovations algorithm* uses these. It does not assume stationarity.

Theorem 8.4 (The Innovations Algorithm). *If $\{X_t\}$ has zero-mean and $\mathbb{E}[X_i X_j] = \kappa(i,j)$, where the $n \times n$ matrix K with entries $K_{ij} = \kappa(i,j)$ is non-singular, then*

$$\widehat{X}_{n+1} = \begin{cases} 0 & \text{if } n = 0, \\ \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & \text{if } n \geq 1, \end{cases} \tag{8.18}$$

where:

$$\begin{cases} v_0 = \kappa(1,1) \\ \theta_{n,n-k} = v_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right) \quad k = 0,\ldots,n-1, \\ v_n = \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \end{cases} \tag{8.19}$$

Proof Taking the inner product of Equation (8.18) with $X_{k+1} - \widehat{X}_{k+1}$ and using orthogonality, it follows that

$$\begin{aligned} \langle \widehat{X}_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle &= \sum_{j=1}^{n} \theta_{n,j} \langle X_{n+1-j} - \widehat{X}_{n+1-j}, X_{k+1} - \widehat{X}_{k+1} \rangle \\ &= \theta_{n,n-k} \|X_{k+1} - \widehat{X}_{k+1}\|^2 = \theta_{n,n-k} v_k, \end{aligned}$$

from which

$$\theta_{n,n-k} = \frac{1}{v_k} \langle X_{n+1}, X_{k+1} - \widehat{X}_{k+1} \rangle.$$

From Equation (8.18), with n replaced by k,

$$\begin{aligned} \theta_{n,n-k} &= v_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \widehat{X}_{j+1} \rangle\right) \\ &= v_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right). \end{aligned}$$

The form of v_n follows immediately from $\|X_n - \widehat{X}_n\|^2 = \|X_n\|^2 - \|\widehat{X}_n\|^2$. □

Prediction of an ARMA Process Using the Innovations Algorithm The innovations algorithm can be applied *directly* to the causal ARMA process

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

but the calculations are simplified considerably by using the following transformation of the process:

$$\begin{cases} W_t = \sigma^{-1} X_t & t = 1,\ldots,m, \\ W_t = \sigma^{-1}\phi(B)X_t, & t > m, \end{cases} \tag{8.20}$$

where $m = \max(p, q)$. It follows that, for $t > m$, W is the MA(q) process $W = \theta(B)\frac{Z}{\sigma}$. It is clear that:

$$\mathcal{H}_n := \overline{\text{spa}}\{X_1, X_2, \ldots, X_n\} = \overline{\text{spa}}\{W_1, W_2, \ldots, W_n\}.$$

As usual, set $\widehat{X}_{n+1} = P_{\mathcal{H}_n}(X_{n+1})$ and $\widehat{W}_{n+1} = P_{\mathcal{H}_n}(W_{n+1})$. It follows directly that

$$\begin{cases} \widehat{W}_t = \sigma^{-1}\widehat{X}_t & t = 1, \ldots, m, \\ \widehat{W}_t = \sigma^{-1}[\widehat{X}_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}] & t > m. \end{cases}$$

From this,

$$X_t - \widehat{X}_t = \sigma(W_t - \widehat{W}_t) \quad \text{for all } t \geq 1.$$

The idea is now to apply the innovations algorithm to $\{W_t\}$. A straightforward computation gives:

$$\kappa_W(i,j) = \begin{cases} \sigma^{-2}\gamma_X(i-j) & 1 \leq i, j \leq m \\ \sigma^{-2}\left[\gamma_X(i-j) - \sum_{r=1}^{p} \phi_r \gamma_X(r - |i-j|)\right] & \min(i,j) \leq m < \max(i,j) \leq 2m \\ \sum_{r=0}^{q} \theta_r \theta_{r+|i-j|}, & \max(i,j) > m, \\ 0 & \text{otherwise}, \end{cases} \tag{8.21}$$

where $\theta_0 = 1$ and $\theta_j = 0$ for $j > q$. The innovations algorithm, applied to W_t gives the coefficients θ_{nj}, from which it follows that:

$$\widehat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & 1 \leq n < m, \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & n \geq m. \end{cases} \tag{8.22}$$

8.2.1 Multivariate Innovations Algorithm

Proposition 8.5 (Multivariate Innovations Algorithm). *Let $\{\underline{X}_t : t \in \mathbb{Z}\}$ be an m dimensional time series with mean $\mathbb{E}[\underline{X}_t] = \underline{0}$ and covariance function $K(i,j) = \mathbb{E}\left[\underline{X}_i \underline{X}_j^t\right]$. Suppose the covariance function of vect$(\underline{X}_1, \ldots, \underline{X}_n)$ is non singular for every $n \geq 1$. Then the one-step predictors $\widehat{\underline{X}}_{n+1} : n \geq 0$ and their prediction error covariance matrices $V_n : n \geq 1$ are given by:*

$$\widehat{\underline{X}_t} = \begin{cases} \underline{0} & n = 0 \\ \sum_{j=1}^{n} \Theta_{n,j}\left(\underline{X}_{n+1-j} - \widehat{\underline{X}}_{n+1-j}\right) & n \geq 1 \end{cases} \tag{8.23}$$

and

$$\begin{cases} V_0 = K(1,1) \\ \Theta_{n,n-k} = \left(K(n+1,k+1) - \sum_{j=0}^{k-1} \Theta_{n,n-j} V_j \Theta_{k,k-j}^t\right) V_k^{-1} & k = 0, 1, \ldots, n-1 \\ V_n = K(n+1,n+1) - \sum_{j=0}^{n-1} \Theta_{n,n-j} V_j \Theta_{n,n-j}^t \end{cases} \tag{8.24}$$

The recursions are solved in the order:

$$V_0; \Theta_{11}, V_1; \Theta_{22}, \Theta_{21}, V_2; \Theta_{33}, \Theta_{32}, \Theta_{31}, V_3;$$

Proof For $i < j$, $\underline{X}_i - \widehat{\underline{X}}_i \in \mathcal{F}_{j-1}^{(X)}$ and since each component of $\underline{X}_j - \widehat{\underline{X}}_j$ is orthogonal to $\mathcal{F}_{j-1}^{(X)}$ by the prediction equation, it follows that

$$\underline{X}_i - \widehat{\underline{X}}_i \perp \underline{X}_j - \widehat{\underline{X}}_j \qquad i \neq j.$$

Post multiplying both sides of Equation (8.23) by $\underline{X}_{k+1}^t - \widehat{\underline{X}}_{k+1}^t$ for $0 \leq k \leq n$ and taking expectations gives:

$$\mathbb{E}\left[\widehat{\underline{X}}_{n+1}\left(\underline{X}_{k+1} - \widehat{\underline{X}}_{k+1}\right)^t\right] = \Theta_{n,n-k}V_k$$

from which it follows, by orthogonality, that

$$\mathbb{E}\left[\underline{X}_{n+1}\left(\underline{X}_{k+1} - \widehat{\underline{X}}_{k+1}\right)^t\right] = \mathbb{E}\left[\widehat{\underline{X}}_{n+1}\left(\underline{X}_{k+1} - \widehat{\underline{X}}_{k+1}\right)^t\right] = \Theta_{n,n-k}V_k.$$

It follows that

$$\Theta_{n,n-k}V_k = K(n+1,k+1) - \sum_{j=0}^{k-1}\mathbb{E}\left[\underline{X}_{n+1}\left(\underline{X}_{j+1} - \widehat{\underline{X}}_{j+1}\right)^t\right]\Theta_{k,k-j}^t$$

from which

$$\Theta_{n,n-k}V_k = K(n+1,k+1) - \sum_{j=0}^{k-1}\Theta_{n,n-j}V_j\Theta_{k,k-j}^t.$$

Nonsingularity of the covariance matrix of $\text{vect}(\underline{X}_1 \ldots, \underline{X}_n)$ gives that V_k is non singular, from which

$$\Theta_{n,n-k} = \left(K(n+1,k+1) - \sum_{j=0}^{k-1}\Theta_{n,n-j}V_j\Theta_{k,k-j}^t\right)V_k^{-1}.$$

Finally, since

$$\underline{X}_{n+1} = \underline{X}_{n+1} - \widehat{\underline{X}}_{n+1} + \sum_{j=0}^{n-1}\Theta_{n,n-j}\left(\underline{X}_{j+1} - \widehat{\underline{X}}_{j+1}\right),$$

it follows, by orthogonality, that

$$K(n+1,n+1) = V_n + \sum_{j=0}^{n-1}\Theta_{n,n-j}V_j\Theta_{n,n-j}^t$$

as required. □

Multivariate Innovations and Recursive Prediction of the VARMA(p,q) Process Let $\{\underline{X}_t : t \in \mathbb{Z}\}$ be an m dimensional causal VARMA(p,q) process,

$$\Phi(B)\underline{X}_t = \Theta(B)\underline{Z}_t \qquad \{\underline{Z}_t\} \sim \text{WN}(\underline{0}, \Sigma)$$

where, using I to denote the $m \times m$ identity matrix,

$$\Phi(B) = I - \Phi_1 B - \ldots - \Phi_p B^p, \qquad \Theta(B) = I + \Theta_1 B + \ldots + \Theta_q B^q, \quad \det(\Sigma) \neq 0$$

As with the univariate ARMA, substantial computational savings can be made by applying the innovations algorithm to the transformed process:

$$\begin{cases} \underline{W}_t = \underline{X}_t & t = 1, \ldots, \max(p,q) \\ \underline{W}_t = \Phi(B)\underline{X}_t & t \geq \max(p,q) + 1 \end{cases}$$

Let Γ denote the covariance function for $\{\underline{X}_t : t \in \mathbb{Z}\}$ and let K denote the covariance function for $\underline{W}$. Let $l = \max(p,q)$. Then

$$K(i,j) := \mathbb{E}\left[\underline{W}_i \underline{W}_j^t\right] = \begin{cases} \Gamma(i-j) & 1 \leq i \leq j \leq l \\ \Gamma(i-j) - \sum_{r=1}^{p} \Phi_r \Gamma(i+r-j) & 1 \leq i \leq l < j \leq 2l \\ \sum_{r=0}^{q} \Theta_r \Sigma \Theta_{r+j-i}^t & l < i \leq j \leq i+q \\ 0 & l < i, \quad i+q < j \\ K^t(i,j) & j < i \end{cases} \tag{8.25}$$

By convention, $\Theta_j \equiv 0_{m \times m}$ for $j > q$. The covariance matrix for this process is zero when $|i-j| > q$. The argument for the univariate setting carries over almost unaltered in the multivariate setting to give:

$$\widehat{\underline{X}}_{n+1} = \begin{cases} \sum_{j=1}^{n-1} \Theta_{nj} \left(\underline{X}_{n+1-j} - \widehat{\underline{X}}_{n+1-j}\right) & 1 \leq n \leq l \\ \Phi_1 \underline{X}_n + \ldots + \Phi_p \underline{X}_{n+1-p} + \sum_{j=1}^{q} \Theta_{nj} \left(\underline{X}_{n+1-j} - \widehat{\underline{X}}_{n+1-j}\right) & n > l \end{cases}$$
$$V_n = \mathbb{E}[(\underline{X}_{n+1} - \widehat{\underline{X}}_{n+1})(\underline{X}_{n+1} - \widehat{\underline{X}}_{n+1})^t]$$

where $\Theta_{nj} : j = 1, \ldots, n$ and V_n are found in Equation (8.24) with $K(i,j)$ defined by Equation (8.25).

For a univariate ARMA, the coefficients $\theta_{nj} : j = 1, \ldots, q$ do not depend on the white noise variance σ^2. In the multivariate case, though, the coefficients do depend on Σ.

When $\{\underline{X}_t : t \in \mathbb{Z}\}$, the differences $\underline{X}_{n+1} - \widehat{\underline{X}}_{n+1}$ satisfy:

$$\mathbb{E}\left[\left(\underline{X}_{n+1} - \widehat{\underline{X}}_{n+1} - \underline{Z}_{n+1}\right)\left(\underline{X}_{n+1} - \widehat{\underline{X}}_{n+1} - \underline{Z}_{n+1}\right)^t\right] \overset{n \to +\infty}{\longrightarrow} 0.$$

It is left as an exercise to show that

$$\lim_{n \to +\infty} \Theta_{nj} = \Theta_j \qquad j = 1, \ldots, q, \qquad \lim_{n \to +\infty} V_n = \Sigma.$$

VARMA(p,q) h step predictors Once $\widehat{\underline{X}}_1, \ldots, \widehat{\underline{X}}_n$ have been computed, it is straightforward to see that the h step predictor $P_{\mathcal{F}_{1:n}^{(X)}} \underline{X}_{n+h}$ satisfies:

$$P_{\mathcal{F}_{1:n}^{(X)}} \underline{X}_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \Theta_{n+h-1} \left(\underline{X}_{n+h-j} - \widehat{\underline{X}}_{n+h-j} \right) & 1 \leq h \leq l-n \\ \sum_{i=1}^{p} \Phi_i P_{\mathcal{F}_{1:n}^{(X)}} \underline{X}_{n+h-i} + \sum_{h \leq j \leq q} \Theta_{n+h-1,j} \left(\underline{X}_{n+h-j} - \widehat{\underline{X}}_{n+h-j} \right) & h \geq l-n+1 \end{cases} \tag{8.26}$$

where, for fixed n, the predictors $P_{\mathcal{F}_{1:n}^{(X)}} X_{n+j} : j = 1, 2, 3, \ldots$ are determined recursively from Equation (8.26).

Let $\underline{g}(h) := P_{\mathcal{F}_{1:n}^{(X)}} \underline{X}_{n+h}$, then $\underline{g}(h)$ satisfies:

$$\begin{cases} \underline{g}(h) - \Phi_1 \underline{g}(h-1) - \ldots - \Phi_p \underline{g}(h-p) = \underline{0} & h > q \\ \underline{g}(q-i) = P_{\mathcal{F}_{1:n}^{(X)}} \underline{X}_{n+q-i} & i = 0, 1, \ldots, p-1. \end{cases} \tag{8.27}$$

8.3 Large numbers of observations

Assume now that we at time 0 have observed $X_{-n+1} \ldots, X_0$ and want to predict X_1 or, more generally X_h for $h \geq 1$. None of the approaches outlined so far give good results if n is very large.

Consider a predictor of X_h based on X_k for $k \leq 0$;

$$P_h(X_h) = P_{\overline{\text{spa}}\{X_k,\, k \leq 0\}} X_h.$$

Any linear predictor $P_h(X_h)$ based on $P_{\overline{\text{spa}}\{X_k, k \leq 0\}} X_h$ is of the form:

$$\widehat{X}_h = \sum_{j=0}^{\infty} \alpha_j X_{-j}. \tag{8.28}$$

It follows from Equation (2.8) that a linear predictor is $P_h(X_h)$ is determined by the relations:

$$\mathbf{C}(P_h(X_h), X_{-i}) = \mathbf{C}(X_h, X_{-i}), \quad i = 0, 1, \ldots,$$

which gives

$$\sum_{j=0}^{\infty} \gamma_X(i-j)\alpha_j = \gamma_X(h+i), \quad i = 0, 1, \ldots. \tag{8.29}$$

This set of equations determines the set of coefficients and the result, $P_h(X_h)$ in Equation (8.28) is the linear predictor provided the resulting series converges; otherwise the expression is not well defined.

Example 8.2 (MA(1) process)**.**

Let $\{X_t\}$ be an MA(1) process, i.e.

$$X_t = Z_t + \theta Z_{t-1} \quad \{Z_t\} \sim \text{WN}(0, \sigma^2).$$

For the MA(1) process, the ACVF is:

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0, \\ \theta\sigma^2 & \text{if } |h| = 1, \\ 0 & \text{if } |h| > 1. \end{cases}$$

The coefficients for the linear predictor are obtained from Equation (8.29). For $h \geq 2$, the equation is satisfied by $\alpha_j \equiv 0$ and the linear predictor is: $P_h(X_h) = 0$. Clearly, for $h \geq 2$, X_h is uncorrelated with any of the observations.

For $h = 1$, Equation (8.29) reduces to

$$\begin{cases} \alpha_0(1+\theta^2) + \alpha_1\theta = \theta \\ \alpha_{i-1}\theta + \alpha_i(1+\theta^2) + \alpha_{i+1}\theta = 0 \quad i = 1, 2, \ldots \end{cases} \tag{8.30}$$

A standard way of solving solutions of this type is to consider the *generating polynomial* $f(z) = \sum_{i=0}^{\infty} \alpha_i z^i$ From Equation (8.30), setting $\alpha_{-1} = 0$,

$$\theta \sum_{i=0}^{\infty} \alpha_{i-1} z^i + (1+\theta^2) \sum_{i=0}^{\infty} \alpha_i z^i + \theta \sum_{i=0}^{\infty} \alpha_{i+1} z^i = \alpha_0(1+\theta^2) + \alpha_1\theta = \theta.$$

From this,

$$\theta z f(z) + (1+\theta^2) f(z) + \frac{\theta}{z}(f(z) - \alpha_0) = \theta$$

from which

$$f(z) = \frac{\theta z + \alpha_0}{\theta z^2 + (1+\theta^2)z + \theta} = \frac{z + \frac{\alpha_0}{\theta}}{(z+\theta)(z+\frac{1}{\theta})}. \tag{8.31}$$

Any $\alpha_0 \in \mathbb{R}$ gives a solution; the coefficients $(\alpha_i)_{i \geq 1}$ can be obtained (in terms of α_0) by taking the coefficients polynomial expansion.

There are additional restrictions, though: convergence requires the additional assumption that $\sum_{i=0}^{\infty} |\alpha_i| < +\infty$. This requires that the polynomial $f(z)$ has no singularities in the unit ball $\{z \in \mathbb{C} : |z| \leq 1\}$

From the equation for $f(z)$, there are three cases: $|\theta| > 1$, $|\theta| < 1$ and $|\theta| = 1$.

Case 1: $|\theta| > 1$: If $|\theta| > 1$, then there is no singularity for $|z| \leq 1$ if and only if $z + \frac{\alpha_0}{\theta} = z + \theta$ for all $z \in \mathbb{C}$ and hence $\alpha_0 = \theta^2$. Therefore α_0 is uniquely determined,

$$f(z) = \theta \sum_{j=0}^{\infty} \left(\frac{-1}{\theta}\right)^j z^j$$

giving $\alpha_j = (-1)^j \frac{1}{\theta^{j-1}}$ for $j \geq 0$.

Case 2: $|\theta| < 1$**:** If $|\theta| < 1$, then there is no singularity for $|z| \leq 1$ if and only if $\alpha_0 = 1$. Therefore α_0 is uniquely deterimined,

$$f(z) = \theta \sum_{j=0}^{\infty} (-1)^j \theta^j z^j$$

giving $\alpha_j = (-1)^j \theta^{j+1}$ for $j \geq 0$.

Case 3: $|\theta| = 1$**:** If $|\theta| = 1$, then

$$f(z) = \frac{\alpha_0 + \theta z}{(z+\theta)^2}$$

so that $\alpha_0 = 1$ and

$$f(z) = \theta \frac{1}{z+\theta} = \begin{cases} \frac{1}{1-z} & \theta = -1 \\ \frac{1}{1+z} & \theta = 1. \end{cases}$$

The expansion is not well defined when $\theta = \pm 1$. Hence there are no solutions to the linear prediction problem of this form in this case. □

Consider again $|\theta| < 1$. The linear predictor is:

$$\widehat{X}_1 = \sum_{j=0}^{\infty} -(-\theta)^{j+1} X_{-j}.$$

Recall that $X_1 = Z_1 + \theta Z_0$. Furthermore, an MA(1) process with $|\theta| < 1$ is invertible and Z_0 may be written as:

$$Z_0 = X_0 - \theta X_{-1} + \theta^2 X_{-2} - \theta^3 X_{-3} + \ldots.$$

From this, it follows that:

$$Z_0 = P_{\overline{\text{spa}}\{X_0, X_{-1}, \ldots\}}(Z_0)$$

Furthermore, $Z_1 \perp \overline{\text{spa}}\{X_0, X_{-1}, \ldots\}$, from which

$$\widehat{X}_1 = P_{\overline{\text{spa}}\{X_0, X_{-1}, \ldots\}}(Z_1) + \theta P_{\overline{\text{spa}}\{X_0, X_{-1}, \ldots\}}(Z_0) = \theta Z_0.$$

It follows that $\widehat{X}_1 = \theta Z_0$.

The natural question arises: When does there exists a predictor of the assumed form? This cannot be answered in general, but the following theorem by Rozanov gives a partial answer.

Theorem 8.6. *Let $\{X_t\}$ be a stationary time series with spectral density $f(\cdot)$. Any $Y \in \overline{spa}\{X_t,\ t \in \mathbb{Z}\}$ can be expressed in the form $Y = \sum_{-\infty}^{\infty} \psi_t X_t$ if and only if*

$$0 < c_1 \leq f(\lambda) \leq c_2 < \infty \quad \textit{for (almost) all } \lambda \in [-\pi, \pi].$$

Proof Omitted □

Example 8.3.

For an MA(1) process,

$$f(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-ih\lambda}\gamma(h) = \frac{(1+\theta^2)\sigma^2}{2\pi} + \frac{\theta\sigma^2}{2\pi}\left(e^{i\lambda} + e^{-i\lambda}\right) = \frac{\sigma^2\left(1+\theta^2+2\theta\cos(\lambda)\right)}{2\pi}, \quad -\pi \leq \lambda \leq \pi.$$

The upper bound is straightforward;

$$\max_{-\pi\leq\lambda\leq\pi} |f(\lambda)| \leq \frac{1}{2\pi}|\gamma(0)| + \frac{1}{\pi}|\gamma(1)| = \frac{\sigma^2}{2\pi}(1+\theta^2) + \frac{\sigma^2}{\pi}|\theta| < +\infty.$$

From the equation for $f(\lambda)$, it follows that $f(0) = 0$ if $\theta = -1$ and $f(\pi) = f(-\pi) = 0$ if $\theta = 1$, while

$$f(\lambda) \geq \frac{\sigma^2\left(1+\theta^2-2|\theta|\right)}{2\pi} = \frac{\sigma^2\left(1-|\theta|\right)^2}{2\pi} > 0, \quad -\pi \leq \lambda \leq \pi$$

for all $|\theta| \neq 1$.

In general, predictors based on infinitely many observations are best expressed in terms of spectral properties of the underlying time series. We will return to this later. Here we only give a preliminary version of Kolmogorov's formula for the *one-step mean-square prediction error*

$$v_\infty = \mathbb{E}\left[(\widehat{X}_{n+1} - X_{n+1})^2\right] \text{ where } \widehat{X}_{n+1} = P_{\overline{\text{spa}}\{X_t,\, t\leq n\}}(X_{n+1}).$$

Theorem 8.7 (Kolmogorov's formula)**.** *Let $\{X_t\}$ be a zero-mean stationary time series with spectral density f such that there exist two constants (c_1, c_2) satisfying*

$$0 < c_1 \leq c_2 < +\infty$$

such that

$$0 < c_1 \leq f(\lambda) \leq c_2 \qquad \textit{for (almost) all} \qquad \lambda \in [-\pi, \pi].$$

The one-step mean-square prediction error is:

$$v_\infty = 2\pi \exp\left\{\tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda\right\}.$$

Proof Firstly, using the expansion

$$\ln(1-z) = \sum_{j=1}^{\infty} \frac{z}{j} \qquad |z| < 1,$$

together with $\int_{-\pi}^{\pi} e^{ik\lambda} d\lambda = 0$ for $k \neq 0$, it follows that for $|a| < 1$,

$$\begin{aligned}\int_{-\pi}^{\pi} \ln|1 - ae^{-i\lambda}|^2 d\lambda &= \int_{-\pi}^{\pi} \left(\ln(1 - ae^{-i\lambda}) + \ln(1 - \overline{a}e^{i\lambda})\right) d\lambda \\ &= -\int_{-\pi}^{\pi} \left(\sum_{j=1}^{\infty} \frac{a^j e^{-ij\lambda}}{j} + \sum_{j=1}^{\infty} \frac{\overline{a}^j e^{ij\lambda}}{j}\right) d\lambda = 0.\end{aligned}$$

Let $\{X_t\}$ be an AR(p) process satisfying $\phi(B)X_t = Z_t$, where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \qquad |z| \leq 1$$

then $\{X_t\}$ has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left|1 - \sum_{j=1}^{p} \phi_j e^{-ij\lambda}\right|^2} = \frac{\sigma^2}{2\pi} \prod_{j=1}^{p} \frac{1}{|1 - a_j e^{-i\lambda}|^2}$$

for a collection of numbers $a_1, \ldots, a_p$ satisfying $|a_j| < 1$ for each $j \in \{1, \ldots, p\}$. It follows that

$$\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda = 2\pi \ln \frac{\sigma^2}{2\pi} - \sum_{j=1}^{p} \int_{-\pi}^{\pi} \ln|1 - a_j e^{-i\lambda}|^2 d\lambda = 2\pi \ln \frac{\sigma^2}{2\pi}.$$

This establishes Kolmogorov's formula for the causal AR process. The result, in the general case, follows from the result, Theorem 3.8, that a stationary linear process can be approximated arbitrarily closely by AR(p) processes. □