2.5: Forecasting Stationary Time Series

Taylor

University of Virginia

Motivation

Goal of this section, assuming we know γ and μ, find a linear combination of $1, X_1, \ldots, X_n$ that does the "best" job of forecasting X_{n+h}, for some $h \in \mathbb{Z}^+$.

Notation

The best linear predictor is denoted $P_n X_{n+h}$. We need to figure out the the coefficients $a_0, a_1, \ldots, a_n$ that minimize the Mean Square Error (MSE)

$$S(a_0, \ldots, a_n) = E[(X_{n+h} - \{a_0 + a_1 X_n \cdots + a_n X_1\})^2].$$

Clearly $S \geq 0$. To minimize we can take derivatives with respect to coefficients and set equal to 0.

$$E[(X_{n+h} - \{a_0 + a_1 X_n \cdots + a_n X_1\})] = 0$$

and

$$E[(X_{n+h} - \{a_0 + a_1 X_n + \cdots + a_n X_1\})X_{n+1-j}] = 0, j = 1, \ldots, n$$

Motivation

Re-arranging and keeping in mind that $j = 1, \ldots, n$

$$E[(X_{n+h} - \{a_0 + a_1 X_n \cdots + a_n X_1\})] = 0$$
$$E[(X_{n+h} - \{a_0 + a_1 X_n + \cdots + a_n X_1\})X_{n+1-j}] = 0$$

becomes

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

$$E[X_{n+h} X_{n+1-j}] = a_0 \mu + \sum_{i=1}^{n} a_i E[X_{n+1-i} X_{n+1-j}]$$

or

$$E[X_{n+h} X_{n+1-j}] - \mu^2 = a_0 \mu + \sum_{i=1}^{n} a_i E[X_{n+1-i} X_{n+1-j}] - \mu^2$$

Motivation

Re-writing that last line

$$E[X_{n+h}X_{n+1-j}] - \mu^2 = a_0\mu + \sum_{i=1}^{n} a_i E[X_{n+1-i}X_{n+1-j}] - \mu^2$$

$$\begin{aligned}
E[X_{n+h}X_{n+1-j}] - \mu^2 &= a_0\mu + \sum_{i=1}^{n} a_i E[X_{n+1-i}X_{n+1-j}] - \mu^2 \\
&= \mu^2\left(1 - \sum_{i=1}^{n} a_i\right) + \sum_{i=1}^{n} a_i E[X_{n+1-i}X_{n+1-j}] - \mu^2 \\
&= -\mu^2 \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} a_i E[X_{n+1-i}X_{n+1-j}] \\
&= \sum_{i=1}^{n} a_i \gamma(i-j)
\end{aligned}$$

The prediction equations

Writing everything with matrices and vectors we have

The Prediction Equations

P_n is determined by the $(a_0, a_1, \ldots, a_n)$ that satisify

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

and

$$\Gamma_n \mathbf{a}_n = \gamma_n(h)$$

where $\gamma_n(h) = (\gamma(h), \ldots, \gamma(h+n-1))'$ and $\Gamma_n = [\gamma(i-j)]_{i,j=1}^{n}$

The Best MSE

Once we have $\mathbf{a}_n$ we can find the lowest possible MSE

$$
\begin{aligned}
E[(X_{n+h} - P_n X_{n+h})^2] &= \text{Var}[X_{n+h} - a_0 - a_1 X_n \cdots - a_n X_1] \\
&= \text{Var}[X_{n+h} - a_1 X_n \cdots - a_n X_1] \\
&= \text{Cov}(X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}, X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}) \\
&= \gamma(0) - 2\sum_{i=1}^{n} a_i \gamma(h+i-1) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(j-i) \\
&= \gamma(0) - 2\mathbf{a}_n' \gamma_n(h) + \mathbf{a}_n' \Gamma_n \mathbf{a}_n \\
&= \gamma(0) - 2\mathbf{a}_n' \gamma_n(h) + \mathbf{a}_n' \gamma_n(h) \\
&= \gamma(0) - \mathbf{a}_n' \gamma_n(h)
\end{aligned}
$$

Example

Let's derive the h-step ahead prediction operator for a causal, mean-zero AR(1) model $X_t = \phi X_{t-1} + Z_t$.

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

and

$$\Gamma_n \mathbf{a}_n = \gamma_n(h)$$

becomes (after dividing both sides by $\gamma(0)$)

$$\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi^h \\ \phi^{h+1} \\ \vdots \\ \phi^{h+n-1} \end{bmatrix} \tag{1}$$

solution: $\mathbf{a}_n = (\phi^h, 0, \ldots, 0)'$. What happens as you try to predict farther out?

Example

What about if we include a mean? $X_t = \mu + \phi(X_{t-1} - \mu) + Z_t$ where the errors are white noise with variance σ^2. Now the intercept $a_0 \neq 0$.

We have this same set of equations again (same as last slide)

$$\Gamma_n \mathbf{a}_n = \gamma_n(h)$$

but now we have this part (because $\mu \neq 0$)

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

So the solution is the same but now we have an extra α_0 term.

Properties of Prediction Operator

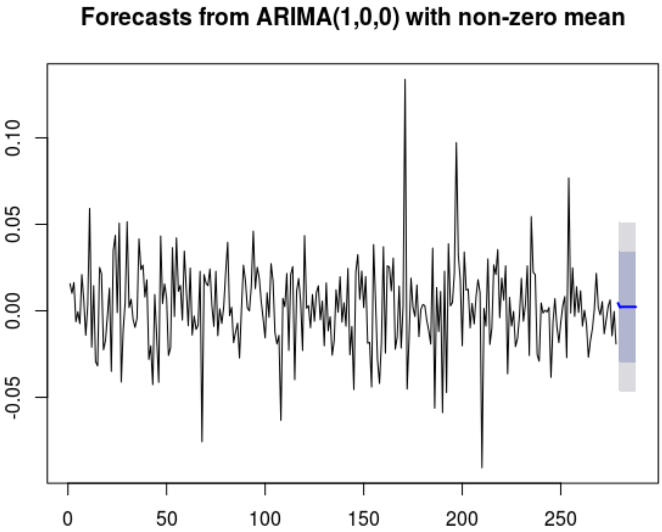
Properties of P_n

- $P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$
- $E[(X_{n+h} - P_n X_{n+h})^2] = \gamma(0) - \mathbf{a}_n' \gamma_n(h)$
- $E(X_{n+h} - P_n X_{n+h}) = 0$
- $E[(X_{n+h} - P_n X_{n+h}) X_j] = 0$ for $j = 1, \ldots, n$
- $P_n X_{n+h}$ is unique.

R Example

```
> fit <- Arima(rets, order=c(1,0,0))
> predict(fit, n.ahead=1) # 0.004414739 = 0.0023 + -0.0998 * (
$pred
Time Series:
Start = 279
End = 279
Frequency = 1
[1] 0.004414739
plot(forecast(fit))
```

R Example


Forecasts from ARIMA(1,0,0) with non-zero mean
0.10
0.05
0.00
-0.05
0
50
100
150
200
250

When $h = 1$: the one-step-ahead predictions

Now let's deal with $h = 1$. Which means X_{n+1} is predicted based on $X_1, \ldots, X_n$:

$$X_{n+1} = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \cdots + \phi_{nn} X_1$$

where

$$\boldsymbol{\phi}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n$$

where $\boldsymbol{\gamma}_n = (\gamma(1), \ldots, \gamma(n))'$ and $\boldsymbol{\phi}_n = (\phi_{n1}, \ldots, \phi_{nn})'$.

We are inverting successively larger and larger matrices! That's very bad.

The Durbin-Levinson algorithm gives us $\boldsymbol{\phi}_n$ in terms of $\boldsymbol{\phi}_{n-1}$.

The Durbin-Levinson Algorithm

Durbin-Levinson

ϕ_n is computed recursively with the following:

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

$$\begin{bmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \phi_{n-1,2} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{n,n} \begin{bmatrix} \phi_{n-1,n-1} \\ \phi_{n-1,n-2} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

and

$$v_n = v_{n-1}[1 - \phi_{nn}^2]$$

The initial points are $\phi_{11} = \rho(1)$ and $v_0 = \gamma(0)$.

The Durbin-Levinson Algorithm

Inductive proof: divide both sides of $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$ by $\gamma(0)$ to get

$$\mathbf{R}_n \boldsymbol{\phi}_n = \boldsymbol{\rho}_n.$$

Clearly this is satisfied for $n = 1$. Suppose it is true for $n = k$ and show that the recursions from the last slide imply that it's true for $n = k + 1$. Let $\boldsymbol{\rho}_k^{(r)} = (\rho(k), \ldots, \rho(1))'$ and $\phi_k^{(r)} = (\phi_{kk}, \ldots, \phi_{k1})'$ ("r" stands for reverse).

$$\begin{aligned}
\mathbf{R}_{k+1}\boldsymbol{\phi}_{k+1} &= \begin{bmatrix} \mathbf{R}_k & \boldsymbol{\rho}_k^{(r)} \\ \boldsymbol{\rho}_k^{(r)'} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_{\boldsymbol{k}} - \phi_{k+1,k+1}\boldsymbol{\phi}_k^{(r)} \\ \phi_{k+1,k+1} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\rho}_k - \phi_{k+1,k+1}\boldsymbol{\rho}_k^{(r)} + \phi_{k+1,k+1}\boldsymbol{\rho}_k^{(r)} \\ \boldsymbol{\rho}_k^{(r)'}\boldsymbol{\phi}_{\boldsymbol{k}} - \phi_{k+1,k+1}\boldsymbol{\rho}_k^{(r)'}\boldsymbol{\phi}_k^{(r)} + \phi_{k+1,k+1} \end{bmatrix} \\
&= \boldsymbol{\rho}_{k+1}
\end{aligned}$$

because $\mathbf{R}_k \boldsymbol{\phi}_k^{(r)} = \boldsymbol{\rho}_k^{(r)}$ too! Last line is a good homework question.

The Durbin-Levinson Algorithm

Note done yet. Need to show the recursions for the mean square errors $v_n = E[(X_{n+1} - P_n X_{n+1})^2]$. Recall we had the initial condition $v_0 = \gamma(0)$, which is true because $P_0 X_1 = E[X_1] = 0$.

$$
\begin{aligned}
v_n &= \gamma(0) - \phi_n' \gamma_n && \text{(2.5 slide 7)} \\
&= \gamma(0) - \begin{bmatrix} \boldsymbol{\phi}_{n-1} - \phi_{nn}\phi_{n-1}^{(r)} \\ \phi_{nn} \end{bmatrix}' \begin{bmatrix} \boldsymbol{\gamma}_{n-1} \\ \gamma(n) \end{bmatrix} && \text{(DL2)} \\
&= \gamma(0) - \phi_{n-1}' \gamma_{n-1} + \phi_{nn}\phi_{n-1}^{(r)\prime}\gamma_{n-1} - \phi_{nn}\gamma(n) \\
&= v_{n-1} + \phi_{nn}\left(\phi_{n-1}^{(r)\prime}\gamma_{n-1} - \gamma(n)\right) && \text{(2.5 slide 7)} \\
&= v_{n-1} + \phi_{nn}\left(\gamma_{n-1}^{(r)\prime}\phi_{n-1} - \gamma(n)\right) && \text{(algebra)} \\
&= v_{n-1} - \phi_{nn}^2\left(\gamma(0) - \phi_{n-1}'\gamma_{n-1}\right) && \text{(DL1)} \\
&= v_{n-1}[1 - \phi_{nn}^2]
\end{aligned}
$$

The Innovations Algorithm

Suppose here that X_t is a zero-mean time series with $E|X_t| < \infty$ for each t, and call $\kappa(i,j) = E[X_i X_j]$. Also denote the 1-step ahead linear predictors as $\hat{X}_n = P_{n-1}X_n$ if $n > 1$ and 0 if $n = 1$. And denote their MSEs as $v_n = E(X_{n+1} - \hat{X}_{n+1})^2$.

The **innovations** are

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

The Innovations Algorithm

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

$$\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{11} & 1 & 0 & \cdots & 0 \\ a_{22} & a_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

or in other words $\mathbf{U} = A_n \mathbf{X}_n$. We also know that $A_n^{-1} = C_n$ is lower-diagonal as well.

The Innovations Algorithm

Our goal here is to write our predictions in terms of our errors.

1. $\mathbf{U}_n = \mathbf{X}_n - \hat{\mathbf{X}}_n$
2. $\mathbf{U}_n = A_n \mathbf{X}_n$
3. $C_n \mathbf{U}_n = \mathbf{X}_n$

so

$$\begin{aligned}
\hat{\mathbf{X}}_n &= \mathbf{X}_n - \mathbf{U}_n \\
&= C_n \mathbf{U}_n - \mathbf{U}_n \\
&= [C_n - I_n]\mathbf{U}_n \\
&= \Theta \mathbf{U}_n
\end{aligned}$$

The Innovations Algorithm

From the last slide we had $\hat{\mathbf{X}}_n = \Theta \mathbf{U}_n$ where

$$\Theta = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}$$

can be re-written as

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

if $n > 0$ and $\hat{X}_1 = 0$. Now we just need recursive formulas for these coefficients

Statement of Algorithm

The Innovations Algorithm

The coefficients $\theta_{n1}, \ldots, \theta_{nn}$ can be computed recursively from the equations

$$v_0 = \kappa(1,1)$$

$$\theta_{n,n-k} = v_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1}\theta_{k,k-j}\theta_{n,n-j}v_j\right), \qquad 0 \le k < n$$

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

No proof given.

Statement of Algorithm

It goes...

1. v_0
2. θ_{11}, v_1
3. $\theta_{22}, \theta_{21}, v_2$
4. $\theta_{33}, \theta_{32}, \theta_{31}, v_3$
5. ...and so on...

Throwback: Example 2.5.5

Derive the recursive predictions for an ARMA(0,1) model

$$X_t = Z_t + \theta_1 Z_{t-1}$$

Everything besides diagonals and off-diagonals of κ are 0.

1. $\kappa(i, i) = \gamma(0) = \sigma^2(1 + \theta^2)$
2. $\kappa(i, i+1) = \theta\sigma^2$

Work on board:

1. $v_0 = (1 + \theta^2)\sigma^2$
2. $v_n = (1 + \theta^2 - v_{n-1}^{-1}\theta^2\sigma^2)\sigma^2$
3. $\theta_{n1} = v_{n-1}^{-1}\theta\sigma^2$
4. and all other coefficients are 0!

Rule of Thumb

General rule of thumb: use Innovations Algo for MA processes. Use Durbin-Levinson for AR processes.

For ARMA processes, we do something a little more clever. For ARMA processes where $p, q \geq 1$, section 3.3 describes how to use the innovations algorithm on a transformed series.