Lecture 11

ARMA Models: Prediction and Forecasting

Reading: Bowerman, O'Connell, and Koehler (2005): Capter 10.3; Cryer and Chen (2008): Chapter 9.1, 9.3, 9.4

MATH 4070: Regression and Time-Series Analysis

Whitney Huang
Clemson University

Agenda

1. **Linear Predictor**

2. **Prediction Equations**

3. **Examples**

Forecasting Stationary Time Series

Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Based on the observed data, $\boldsymbol{X}_n = (X_1, X_2, \cdots, X_n)^T$, we want to forecast X_{n+h} for some h, a positive integer

- Question: What is the best way to do so?
  $\Rightarrow$ Need to decide on what "best" means

Forecasting Stationary Time Series

Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Based on the observed data, $\boldsymbol{X}_n = (X_1, X_2, \cdots, X_n)^T$, we want to forecast X_{n+h} for some h, a positive integer

- Question: What is the best way to do so?
  $\Rightarrow$ Need to decide on what “best” means
- A commonly used metric for describing forecast performance is the mean square prediction error (MSPE):

$$\mathrm{MSPE} = \mathrm{E}\left[\left(X_{n+h} - m_n(\boldsymbol{X}_n)\right)^2\right].$$

  $\Rightarrow$ the best predictor (in terms of MSPE) is

$$m_n(\boldsymbol{X}_n) = \mathbb{E}\left[X_{n+h} | \boldsymbol{X}_n\right],$$

  the conditional expectation of X_{n+h} given $\boldsymbol{X}_n$

Linear Predictor

Calculating $\mathbb{E}\left[X_{n+h} | \boldsymbol{X}_n\right]$ can be difficult in general

- We will restrict to a linear combination of $X_1, X_2, \cdots, X_n$ and a constant ⇒ linear predictor:

$$\begin{aligned} P_n X_{n+h} &= c_0 + c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1 \\ &= c_0 + \sum_{j=1}^{n} c_j X_{n+1-j} \end{aligned}$$

Linear Predictor

Calculating $\mathbb{E}\left[X_{n+h} | \boldsymbol{X}_n\right]$ can be difficult in general

- We will restrict to a linear combination of $X_1, X_2, \cdots, X_n$ and a constant $\Rightarrow$ linear predictor:

$$
\begin{aligned}
P_n X_{n+h} &= c_0 + c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1 \\
&= c_0 + \sum_{j=1}^{n} c_j X_{n+1-j}
\end{aligned}
$$

- We select the coefficients that minimize the h-step-ahead mean squared prediction error:

$$
\mathbb{E}\left(\left[X_{n+h} - P_n X_{n+h}\right]^2\right) = \mathbb{E}\left(X_{n+h} - c_0 - \sum_{j=1}^{n} c_j X_{n+1-j}\right)^2
$$

Linear Predictor

Calculating $\mathbb{E}\left[X_{n+h}|\boldsymbol{X}_n\right]$ can be difficult in general

- We will restrict to a linear combination of $X_1, X_2, \cdots, X_n$ and a constant ⇒ linear predictor:

$$\begin{aligned} P_n X_{n+h} &= c_0 + c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1 \\ &= c_0 + \sum_{j=1}^{n} c_j X_{n+1-j} \end{aligned}$$

- We select the coefficients that minimize the h-step-ahead mean squared prediction error:

$$\mathbb{E}\left(\left[X_{n+h} - P_n X_{n+h}\right]^2\right) = \mathbb{E}\left(X_{n+h} - c_0 - \sum_{j=1}^{n} c_j X_{n+1-j}\right)^2$$

- The best linear predictor is the best predictor if $\{X_t\}$ is Gaussian

How to Determine these Coefficients $\{c_j\}$?

The steps that we are about to follow to calculate the c_j values are the same as you would use for calculating ordinary least squares estimates

1. Take the derivative of the MSPE with respect to each coefficient c_j

How to Determine these Coefficients $\{c_j\}$?

The steps that we are about to follow to calculate the c_j values are the same as you would use for calculating ordinary least squares estimates

1. Take the derivative of the MSPE with respect to each coefficient c_j
2. Set each derivative equal to zero

How to Determine these Coefficients $\{c_j\}$?

The steps that we are about to follow to calculate the c_j values are the same as you would use for calculating ordinary least squares estimates

1. Take the derivative of the MSPE with respect to each coefficient c_j
2. Set each derivative equal to zero
3. Solve with respect to the coefficients

Forecasting Stationary Processes I

For simplicity, let's assume $\mu = 0$ (we can always achieve that by subtracting off μ) so that we don't need the constant term. We have

$$P_n X_{n+h} = c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1.$$

Forecasting Stationary Processes I

For simplicity, let's assume $\mu = 0$ (we can always achieve that by subtracting off μ) so that we don't need the constant term. We have

$$P_n X_{n+h} = c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1.$$

We want the MSPE

$$\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$$

as small as possible.

Forecasting Stationary Processes I

For simplicity, let's assume $\mu = 0$ (we can always achieve that by subtracting off μ) so that we don't need the constant term. We have

$$P_n X_{n+h} = c_1 X_n + c_2 X_{n-1} + \cdots + c_n X_1.$$

We want the MSPE

$$\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$$

as small as possible.

From now on let's definite

$$\mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right] = S(c_1, \cdots, c_n)$$

We are going to take derivative of the $S(c_1, \cdots, c_n)$ with respect to each coefficient c_j

Forecasting Stationary Processes II

S is a quadratic function of $c_1, c_2, \cdots, c_n$, so any minimizing set of c_j's must satisfy these n equations:

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = 0, \quad j = 1, \cdots, n.$$

Since $S(c_1, \cdots, c_n) = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$, we have

Forecasting Stationary Processes II

S is a quadratic function of $c_1, c_2, \cdots, c_n$, so any minimizing set of c_j's must satisfy these n equations:

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = 0, \quad j = 1, \cdots, n.$$

Since $S(c_1, \cdots, c_n) = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$, we have

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = -2\mathbb{E}\left[\left(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}\right) X_{n-j+1}\right] = 0$$

$$\Rightarrow \mathrm{Cov}(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}) = 0, \quad j = 1, \cdots, n$$

Forecasting Stationary Processes II

S is a quadratic function of $c_1, c_2, \cdots, c_n$, so any minimizing set of c_j's must satisfy these n equations:

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = 0, \quad j = 1, \cdots, n.$$

Since $S(c_1, \cdots, c_n) = \mathbb{E}\left[(X_{n+h} - c_1 X_n - c_2 X_{n-1} - \cdots - c_n X_1)^2\right]$, we have

$$\frac{\partial S(c_1, \cdots, c_n)}{\partial c_j} = -2\mathbb{E}\left[\left(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}\right) X_{n-j+1}\right] = 0$$

$$\Rightarrow \mathrm{Cov}(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}) = 0, \quad j = 1, \cdots, n$$

$\Rightarrow$ Prediction error is uncorrelated with all RVs used in corresponding predictor

Forecasting Stationary Processes III

Orthogonality principle:

$$\mathrm{Cov}\left(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}\right) = 0, \quad j = 1, \cdots, n.$$

Forecasting Stationary Processes III

Orthogonality principle:

$$\mathrm{Cov}(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}) = 0, \quad j = 1, \cdots, n.$$

We have

$$\mathrm{Cov}(X_{n+h}, X_{n-j+1}) - \sum_{i=1}^{n} c_i \mathrm{Cov}(X_{n-i+1}, X_{n-j+1}) = 0$$

Forecasting Stationary Processes III

Orthogonality principle:

$$\mathrm{Cov}(X_{n+h} - \sum_{i=1}^{n} c_i X_{n-i+1}, X_{n-j+1}) = 0, \quad j = 1, \cdots, n.$$

We have

$$\mathrm{Cov}(X_{n+h}, X_{n-j+1}) - \sum_{i=1}^{n} c_i \mathrm{Cov}(X_{n-i+1}, X_{n-j+1}) = 0$$

We obtain $\{c_i; i = 1, \cdots, n\}$ by solving the system of linear equations:

$$\left\{\gamma(h+j-1) = \sum_{i=1}^{n} c_i \gamma(i-j) : j = 1, \cdots, n\right\},$$

to find n unknown c_i's

Computing $P_n X_{n+h}$ via Matrix Operations

We can rewrite the system of prediction equations as

$$\boldsymbol{\gamma}_n = \Sigma_n \boldsymbol{c}_n,$$

with $\boldsymbol{\gamma}_n = (\gamma(h), \gamma(h+1), \cdots \gamma(h+n-1))^T$, $\boldsymbol{c}_n = (c_1, c_2, \cdots, c_n)^T$ and

$$\Sigma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}$$

is the covariance matrix of $(X_1, X_2, \cdots, X_n)^T$.

Solving for $\boldsymbol{c}_n$ we have

$$\boldsymbol{c}_n = \Sigma_n^{-1} \boldsymbol{\gamma}_n$$

Properties of the Prediction Errors

The prediction errors are

$$\begin{aligned} U_{n+h} &= X_{n+h} - P_n X_{n+h} \\ &= (X_{n+h} - \mu) - \sum_{j=1}^{n} c_j (X_{n+1-j} - \mu). \end{aligned}$$

It then follows that

- The prediction error has mean zero

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

Properties of the Prediction Errors

The prediction errors are

$$\begin{aligned} U_{n+h} &= X_{n+h} - P_n X_{n+h} \\ &= (X_{n+h} - \mu) - \sum_{j=1}^{n} c_j (X_{n+1-j} - \mu). \end{aligned}$$

It then follows that

- The prediction error has mean zero

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

- The prediction error is uncorrelated with all RVs used in the predictor

$$\operatorname{Cov}(U_{n+h}, X_j) = \operatorname{Cov}(X_{n+h} - P_n X_{n+h}, X_j) = 0, \quad j = 1, \cdots, n$$

The Minimum Mean Squared Prediction Error

We obtain the minimum value of the MSPE by substituting the expression for $\boldsymbol{c}_n$ into $\mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right]$:

$$
\begin{aligned}
\mathrm{MSPE} &= \mathbb{E}\left[(X_{n+h} - P_n X_{n+h})^2\right] \\
&= \mathbb{E}\left[(X_{n+h} - \mu)^2\right] - 2\sum_{j=1}^{n} c_j \mathbb{E}\left[(X_{n+1-j} - \mu)(X_{n+h} - \mu)\right] \\
&+ \mathbb{E}\left[\sum_{j=1}^{n} c_j (X_{n+1-j} - \mu)\right]^2 \\
&= \mathbb{E}\left[(X_{n+h} - \mu)^2\right] - 2\sum_{j=1}^{n} c_j \mathbb{E}\left[(X_{n+1-j} - \mu)(X_{n+h} - \mu)\right] \\
&+ \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k \mathbb{E}\left[(X_{n+1-j} - \mu)(X_{n+1-k} - \mu)\right] \\
&= \gamma(0) - 2\sum_{j=1}^{n} c_j \gamma(h+j-1) + \sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k \gamma(k-j) \\
&= \gamma(0) - 2\boldsymbol{c}_n^T \boldsymbol{\gamma}_n + \boldsymbol{c}_n^T \Sigma_n \boldsymbol{c}_n.
\end{aligned}
$$

The Minimum Mean Squared Prediction Error (Cont'd)

From the previous slide we have

$$\text{MSPE} = \gamma(0) - 2\boldsymbol{c}_n^T\boldsymbol{\gamma}_n + \boldsymbol{c}_n^T\Sigma_n\boldsymbol{c}_n$$

Recall that $\boldsymbol{c}_n = \Sigma_n^{-1}\boldsymbol{\gamma}_n$, therefore we have

$$\begin{aligned}\text{MSPE} &= \gamma(0) - 2\boldsymbol{c}_n^T\boldsymbol{\gamma}_n + \boldsymbol{c}_n^T\Sigma_n\Sigma_n^{-1}\boldsymbol{\gamma}_n \\ &= \gamma(0) - \boldsymbol{c}_n^T\boldsymbol{\gamma}_n \\ &= \gamma(0) - \sum_{j=1}^{n} c_j\gamma(h+j-1).\end{aligned}$$

If $\{X_t\}$ is a Gaussian process then an approximate $100(1-\alpha)$% prediction interval for X_{n+h} is given by

$$P_nX_{n+h} \pm z_{1-\alpha/2}\sqrt{\text{MSPE}}.$$

One-Step Ahead Prediction of AR(1) Process

Consider AR(1) process $X_t = \phi X_{t-1} + Z_t$, where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}(0, 1 - \phi^2)$.

- Since $\mathrm{Var}(X_t) = 1$, $\gamma(h) = \rho(h) = \phi^{|h|}$
- To forecast X_{n+1} based upon $\boldsymbol{X}_n = (X_1, \cdots, X_n)^T$, using best linear predictor $P_n X_{n+1} = \boldsymbol{c}_n^T \boldsymbol{X}_n$, we need to solve $\Sigma_n \boldsymbol{c}_n = \boldsymbol{\gamma}_n$

$$\begin{bmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}$$

$\Rightarrow$ the solution is $\boldsymbol{c}_n = (\phi, 0, \cdots, 0)^T$, yielding

$$P_n X_{n+1} = \boldsymbol{c}_n^T \boldsymbol{X}_n = \phi X_n$$

One-Step Ahead Prediction of AR(1) Process (Cont'd)

- ϕX_n makes intuitive sense as a predictor since

$$X_{n+1} = \phi X_n + Z_{n+1}$$

One-Step Ahead Prediction of AR(1) Process (Cont'd)

- ϕX_n makes intuitive sense as a predictor since

$$X_{n+1} = \phi X_n + Z_{n+1}$$

- Prediction error is $X_{n+1} - \phi X_n = Z_{n+1}$ and

$$\mathrm{Cov}(Z_t, X_{n-j+1}) = 0, \ j = 1, \cdots, n$$

One-Step Ahead Prediction of AR(1) Process (Cont'd)

- ϕX_n makes intuitive sense as a predictor since

$$X_{n+1} = \phi X_n + Z_{n+1}$$

- Prediction error is $X_{n+1} - \phi X_n = Z_{n+1}$ and

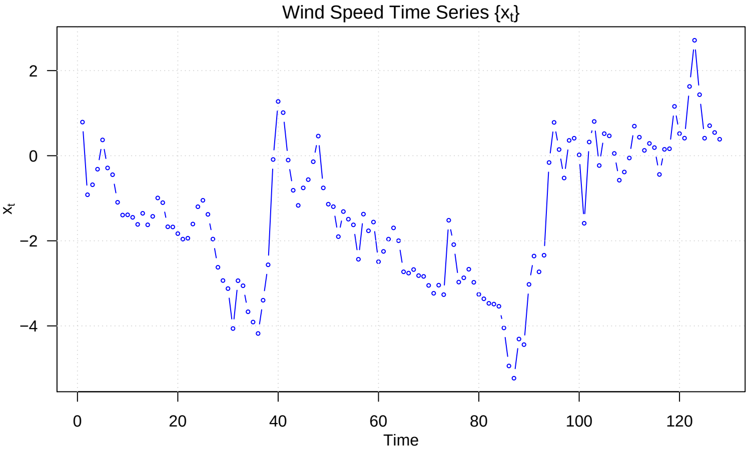
$$\mathrm{Cov}(Z_t, X_{n-j+1}) = 0,\ j = 1, \cdots, n$$

- MSPE is

$$\mathrm{Var}(X_{n+1} - \phi X_n) = \gamma(0) - \boldsymbol{c}_n^T \boldsymbol{\gamma}_n = 1 - \phi^2,$$

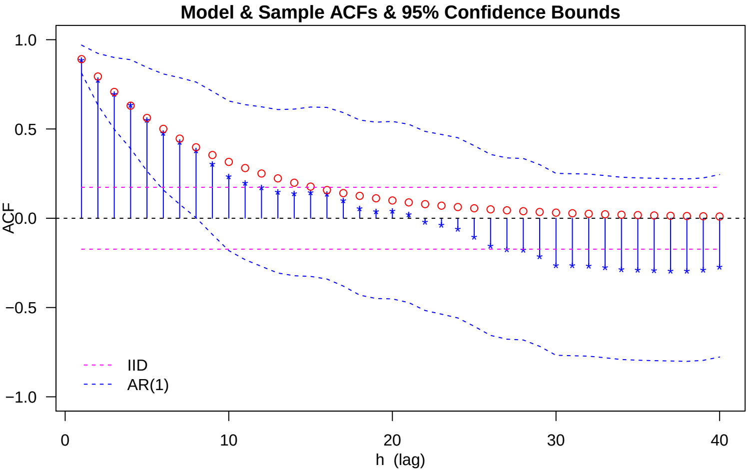
because $\boldsymbol{c}_n = (\phi, 0, \cdots, 0)^T$ and $\boldsymbol{\gamma}_n = (\phi, \phi^2, \cdots, \phi^n)^T$

Wind Speed Time Series Example [Source: UW stat 519 lecture notes by Donald Percival]



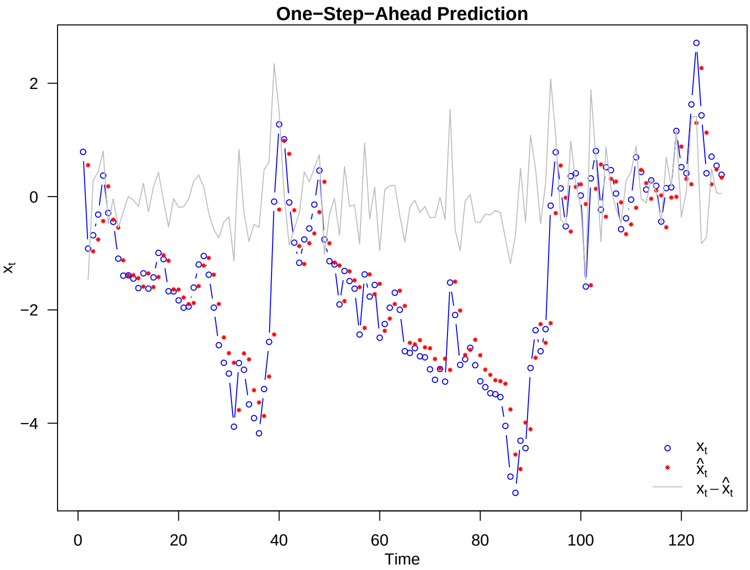
Let's use this series to illustrate forecasting one step ahead

Model & Sample ACFs & 95% Confidence Bounds



The sample ACF indicates compatibility with AR(1) model
$\Rightarrow P_n X_{n+1} = \phi X_n$

One-Step-Ahead Prediction of Wind Speed Series



Predicting "Missing" Values

- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3

Predicting "Missing" Values

- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3

- Solution: To calculate $P_{X_1,X_3}X_2$ we minimize

$$\begin{aligned}\text{MSPE} &= \mathbb{E}\left[(X_2 - P_{X_1,X_3}X_2)^2\right] \\ &= \mathbb{E}\left[(X_2 - c_0 - c_1 X_3 - c_2 X_1)^2\right]\end{aligned}$$

Predicting "Missing" Values

- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3

- Solution: To calculate $P_{X_1,X_3}X_2$ we minimize

$$\begin{aligned}\text{MSPE} &= \mathbb{E}\left[(X_2 - P_{X_1,X_3}X_2)^2\right]\\ &= \mathbb{E}\left[(X_2 - c_0 - c_1X_3 - c_2X_1)^2\right]\end{aligned}$$

- Proceed as for the forecasting case to get the optimal coefficients:

Predicting "Missing" Values

- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3

- Solution: To calculate $P_{X_1,X_3}X_2$ we minimize

$$\begin{aligned} \mathrm{MSPE} &= \mathbb{E}\left[(X_2 - P_{X_1,X_3}X_2)^2\right] \\ &= \mathbb{E}\left[(X_2 - c_0 - c_1 X_3 - c_2 X_1)^2\right] \end{aligned}$$

- Proceed as for the forecasting case to get the optimal coefficients:
  - Calculate derivatives

Predicting "Missing" Values

- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3

- Solution: To calculate $P_{X_1,X_3}X_2$ we minimize

$$\begin{aligned}\mathrm{MSPE} &= \mathbb{E}\left[(X_2 - P_{X_1,X_3}X_2)^2\right] \\ &= \mathbb{E}\left[(X_2 - c_0 - c_1X_3 - c_2X_1)^2\right]\end{aligned}$$

- Proceed as for the forecasting case to get the optimal coefficients:
  - Calculate derivatives
  - Set the derivatives equal to zero

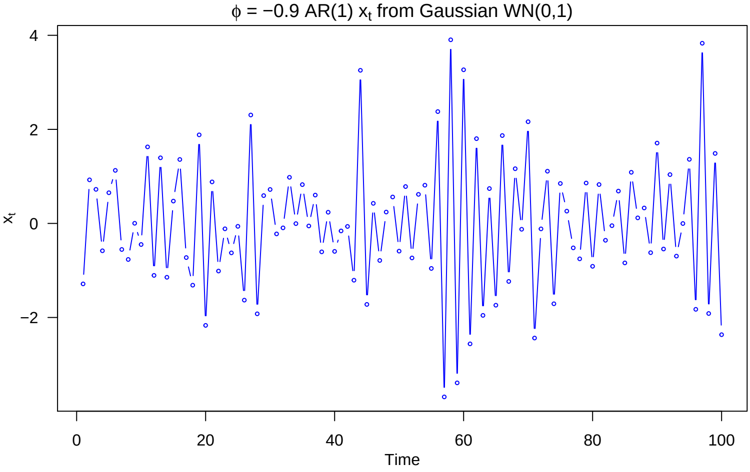
Predicting “Missing” Values

- Let $\{X_t\}$ be a stationary process with mean μ and ACVF $\gamma(\cdot)$. Suppose we know X_1 and X_3, and want to predict X_2 using linear combinations of X_1 and X_3
- Solution: To calculate $P_{X_1,X_3}X_2$ we minimize

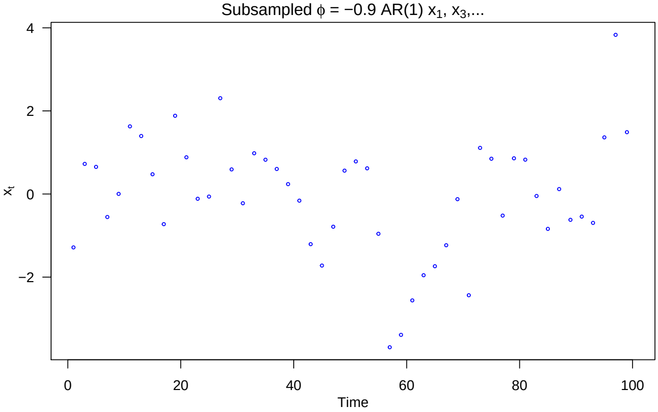
$$\begin{aligned}\text{MSPE} &= \mathbb{E}\left[(X_2 - P_{X_1,X_3}X_2)^2\right]\\ &= \mathbb{E}\left[(X_2 - c_0 - c_1X_3 - c_2X_1)^2\right]\end{aligned}$$

- Proceed as for the forecasting case to get the optimal coefficients:
  - Calculate derivatives
  - Set the derivatives equal to zero
  - Solve the linear system of equation

Another AR(1) Example with $\phi = -0.9$



Subsampled $X_1, X_3, \cdots$ and Removed $X_2, X_4, \cdots$



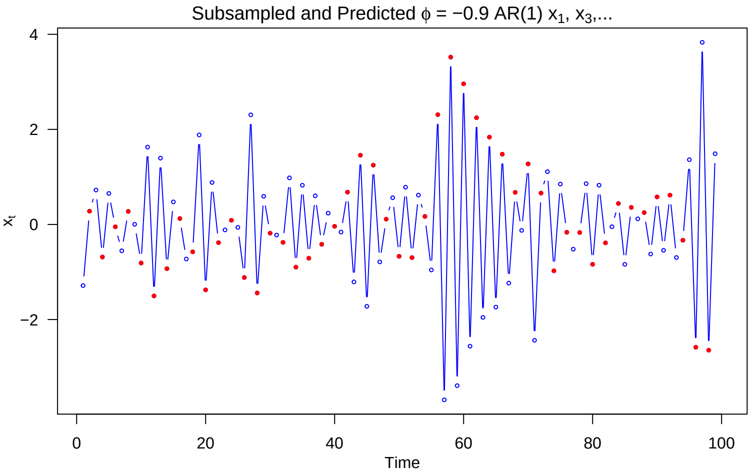
The best linear predictor of X_2 given X_1, X_3 is

$$\hat{X}_2 = \frac{\phi}{1+\phi^2}(X_1 + X_3),$$

and the MSPE is

$$\frac{\sigma^2}{1+\phi^2}$$

Predict $X_2, X_4, \cdots$ Using Best Linear Predictor



Prediction Errors from Best Linear Predictor