Forecasting Stationary Processes: I

- suppose $\{X_t\}$ is a stationary process with mean μ, variance σ_X^2 and ACF $\{\rho(h)\}$, all of which are assumed known
- given $X_1, \ldots, X_n$, suppose we want to forecast (predict) X_{n+h}, where h is a positive integer
- Q: what is the best way to do so?
- to address question, first have to decide on what 'best' means
- to motivate discussion, will focus initially on using just X_n (most recently observed value in time series) to forecast X_{n+h}

Forecasting Stationary Processes: II

- to start with, let's only consider *linear* predictors:
$$\ell(X_n; a_0, a_1) = a_0 + a_1 X_n$$
- suppose we set a_0 and a_1 so that the *mean square error* (MSE)
$$E\{[X_{n+h} - \ell(X_n; a_0, a_1)]^2\}$$
is as small as possible (thus quantifying what 'best' means)
- Cryer & Chan, pp. 219–20: *best linear predictor* is
$$\ell(X_n) \stackrel{\text{def}}{=} \ell(X_n; \mu(1 - \rho(h)), \rho(h)) = \mu + \rho(h)(X_n - \mu)$$
and has a corresponding MSE of
$$E\{[X_{n+h} - \ell(X_n)]^2\} = \sigma_X^2(1 - \rho^2(h))$$
- MSE will be small if $|\rho(h)|$ is close to unity (i.e., strong correlation within a stationary process *helps* prediction)

Forecasting Stationary Processes: III

- more generally (but sticking with MSE as what we mean by 'best'), might consider *best predictor*, say $m(X_n)$, which is the function of X_n (not necessarily linear!) such that

$$E\{[X_{n+h} - m(X_n)]^2\} \text{ is as small as possible}$$

- Cryer & Chan, p. 220, show that $m(X_n) = E\{X_{n+h} \,|\, X_n\}$ with mean square prediction error

$$E\{[X_{n+h} - E\{X_{n+h} \,|\, X_n\}]^2\} \leq E\{[X_{n+h} - \ell(X_n)]^2\}$$

  - $E\{X_{n+h} \,|\, X_n\}$ is conditional expectation of X_{n+h} given X_n
  - given RVs U and V, two basic properties of $E\{U \,|\, V\}$ are
    * $E\{U \,|\, V\} = E\{U\}$ if U and V are independent
    * you can treat V as if it were a constant (it isn't!) – thus, for constants a, b & c, $E\{aU + bV \,|\, V\} = aE\{U \,|\, V\} + bV$, which mimics $E\{aU + bc\} = aE\{U\} + bc$ (see Wikipedia)

Forecasting Stationary Processes: IV

- if we make the additional assumption that $\{X_t\}$ is Gaussian, then $m(X_n)$ and $\ell(X_n)$ are necessarily the same
- in non-Gaussian situations, $m(X_n)$ and $\ell(X_n)$ can differ
- more generally, will be interested in predicting X_{n+h} using X_n, $\ldots$, X_1, leading to best linear predictor

$$\ell(X_n, \ldots, X_1) = a_0 + a_1 X_n + \cdots a_n X_1$$

with a_j's set such that

$$E\{[X_{n+h} - \ell(X_n, \ldots, X_1; a_0, a_1, \ldots, a_n)]^2\}$$

is minimized and with a_j's depending on just μ and $\{\rho(h)\}$

Forecasting Stationary Processes: V

- leads also to best predictor $m(X_n, \ldots, X_1)$, which takes the form
$$m(X_n, \ldots, X_1) = E\{X_{n+h} \,|\, X_n, \ldots, X_1\}$$
and is the same as $\ell(X_n, \ldots, X_1)$ when $\{X_t\}$ is Gaussian
- necessity of knowing joint distributions – and difficulty of computing conditional expectations even when these are known – make $m(X_n, \ldots, X_1)$ much less practical to deal with than $\ell(X_n, \ldots, X_1)$ for non-Gaussian $\{X_t\}$
- focus will thus be on best linear predictors

Forecasting Stationary Processes: VI

- assume $\{X'_t\}$ is stationary with *known* mean μ and *known* ACVF $\{\gamma(h)\}$
- let $X_t = X'_t - \mu$, so that $\{X_t\}$ has mean zero and ACVF $\{\gamma(h)\}$; thus, can assume zero-mean process $\{X_t\}$ and then substitute $X'_t - \mu$ for X_t to handle $\mu \neq 0$ case
- goal: forecast X_{n+h} based upon linear combination of X_n, $X_{n-1}, \ldots, X_1$, say

$$\widehat{X}_{n+h} \stackrel{\text{def}}{=} a_1 X_n + a_2 X_{n-1} + \cdots + a_n X_1,$$

such that its MSE

$$\begin{aligned} E\{(X_{n+h} - \widehat{X}_{n+h})^2\} &= E\{(X_{n+h} - a_1 X_n - \cdots - a_n X_1)^2\} \\ &\stackrel{\text{def}}{=} S(a_1, \ldots, a_n) \end{aligned}$$

is as small as possible

Forecasting Stationary Processes: VII

- S is a quadratic function of $a_1, \ldots, a_n$, so any minimizing set of a_j's must satisfy these n equations:

$$\frac{\partial S(a_1, \ldots, a_n)}{\partial a_j} = 0, \quad j = 1, \ldots, n$$

- since $S(a_1, \ldots, a_n) = E\{(X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1})^2\}$, have

$$\frac{\partial S(a_1, \ldots, a_n)}{\partial a_j} = -2E\{(X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}) X_{n-j+1}\} = 0;$$

i.e., since $E\{X_{n+h} - \sum_i a_i X_{n-i+1}\} = 0$ & $E\{X_{n-j+1}\} = 0$,

$$\operatorname{cov}\{X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}, X_{n-j+1}\} = 0, \quad j = 1, \ldots, n$$

Forecasting Stationary Processes: VIII

- letting $a_1, \ldots, a_n$ now stand for minimizers of S, have
$$\widehat{X}_{n+h} = \sum_{i=1}^{n} a_i X_{n-i+1}$$
- noting $X_{n+h} - \widehat{X}_{n+h}$ is the prediction error and revisiting
$$\text{cov}\left\{X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}, X_{n-j+1}\right\} = 0, \quad j = 1, \ldots, n,$$
we have the celebrated *orthogonality principle*, namely,
$$\text{cov}\left\{X_{n+h} - \widehat{X}_{n+h}, X_{n-j+1}\right\} = 0, \quad j = 1, \ldots, n;$$
i.e., prediction error is uncorrelated with *all* RVs used in corresponding predictor
- note: prediction error $X_{n+h} - \widehat{X}_{n+h}$ has zero mean (why?)

Forecasting Stationary Processes: IX

- returning now to

$$\operatorname{cov}\left\{X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}, X_{n-j+1}\right\} = 0,$$

have

$$\operatorname{cov}\left\{X_{n+h}, X_{n-j+1}\right\} - \sum_{i=1}^{n} a_i \operatorname{cov}\left\{X_{n-i+1}, X_{n-j+1}\right\} = 0,$$

i.e., need to solve system of n equations

$$\sum_{i=1}^{n} a_i \gamma(i-j) = \gamma(h+j-1), \quad j = 1, \ldots, n,$$

to find n unknown a_i's

Forecasting Stationary Processes: X

- equations
$$\sum_{i=1}^{n} a_i\gamma(i-j) = \gamma(h+j-1), \quad j = 1, \ldots, n, \qquad (*)$$
can be written in vector/matrix notation as
$$\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(h), \qquad (**)$$
where Γ_n is an $n \times n$ matrix whose (i, j) entry is $\gamma(i-j)$, while $\boldsymbol{a}_n$ and $\boldsymbol{\gamma}_n(h)$ are n-dimensional column vectors given by
$$\boldsymbol{a}_n = [a_1, \ldots, a_n]' \text{ and } \boldsymbol{\gamma}_n(h) = [\gamma(h), \ldots, \gamma(h+n-1)]'$$
- Γ_n is covariance matrix for $[X_n, X_{n-1}, \ldots, X_1]'$
- $\boldsymbol{\gamma}_n(h)$ has covariances between X_{n+h} and $[X_n, X_{n-1}, \ldots, X_1]'$
- note: dividing both sides of $(*)$ by $\gamma(0)$ shows that $\boldsymbol{a}_n$ does *not* depend on variance of $\{X_t\}$, but rather just on its ACF

Forecasting Stationary Processes: XI

- mean square prediction error given by

$$E\{(X_{n+h} - \widehat{X}_{n+h})^2\}$$

$$= \text{cov}\left\{X_{n+h} - \sum_{i=1}^{n} a_i X_{n-i+1}, X_{n+h} - \sum_{j=1}^{n} a_j X_{n-j+1}\right\}$$

$$= \text{cov}\{X_{n+h}, X_{n+h}\} - 2\sum_{j=1}^{n} a_j \,\text{cov}\{X_{n+h}, X_{n-j+1}\}$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\text{cov}\{X_{n-i+1}, X_{n-j+1}\}$$

$$= \gamma(0) - 2\sum_{j=1}^{n} a_j \gamma(h+j-1) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j)$$

Forecasting Stationary Processes: XII

- noting that

$$\sum_{j=1}^{n} a_j \gamma(h+j-1) = \boldsymbol{a}_n' \boldsymbol{\gamma}_n(h) \text{ and } \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j) = \boldsymbol{a}_n' \Gamma_n \boldsymbol{a}_n,$$

we have

$$\begin{aligned} E\{(X_{n+h} - \widehat{X}_{n+h})^2\} &= \gamma(0) - 2\sum_{j=1}^{n} a_j \gamma(h+j-1) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(i-j) \\ &= \gamma(0) - 2\boldsymbol{a}_n' \boldsymbol{\gamma}_n(h) + \boldsymbol{a}_n' \Gamma_n \boldsymbol{a}_n \\ &= \gamma(0) - \boldsymbol{a}_n' \boldsymbol{\gamma}_n(h) \end{aligned}$$

because $\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(h)$ from $(**)$ implies $\boldsymbol{a}_n' \Gamma_n \boldsymbol{a}_n = \boldsymbol{a}_n' \boldsymbol{\gamma}_n(h)$

Forecasting Stationary Processes: XIII

- letting $\boldsymbol{X}_n \stackrel{\text{def}}{=} [X_n, \ldots, X_1]'$ and recalling that
  1. Γ_n is an $n \times n$ matrix whose (i, j)th element is $\gamma(i - j)$,
  2. $\boldsymbol{\gamma}_n(h) = [\gamma(h), \ldots, \gamma(h + n - 1)]'$ and
  3. $\boldsymbol{a}_n = [a_1, \ldots, a_n]'$ is a solution to $\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(h)$,

  best linear predictor $\widehat{X}_{n+h} = \sum_{i=1}^{n} a_i X_{n-i+1} = \boldsymbol{a}_n' \boldsymbol{X}_n$ of X_{n+h} has the following properties:
  1. $\widehat{X}_{n+h}$ is *unique* even if $\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(h)$ has multiple solutions
  2. $E\{X_{n+h} - \widehat{X}_{n+h}\} = 0$ (prediction error has zero mean)
  3. $\operatorname{var}\{X_{n+h} - \widehat{X}_{n+h}\} = \gamma(0) - \boldsymbol{a}_n' \boldsymbol{\gamma}_n(h)$ (same as mean square prediction error)
  4. $\operatorname{cov}\{X_{n+h} - \widehat{X}_{n+h}, X_{n-j+1}\} = 0$, $j = 1, \ldots, n$ (orthogonality principle, which *uniquely* determines $\widehat{X}_{n+h}$)

One-Step Ahead Prediction of AR(1) Process: I

- as an example, consider AR(1) model $X_t = \phi X_{t-1} + Z_t$, where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}(0, 1-\phi^2)$
- since var $\{X_t\} = 1$ here, ACVF given by $\gamma(h) = \rho(h) = \phi^{|h|}$
- to forecast X_{n+1} based upon $\boldsymbol{X}_n$ using best linear predictor $\widehat{X}_{n+1} = \boldsymbol{a}_n' \boldsymbol{X}_n$, must solve $\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(1)$:

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

- clearly the solution is $\boldsymbol{a}_n = [\phi, 0, \ldots, 0]'$, yielding $\widehat{X}_{n+1} = \boldsymbol{a}_n' \boldsymbol{X}_n = \phi X_n$ (note: we do *not* make use of $X_1, X_2, \ldots X_{n-1}$)

BD: 57, CC: 193

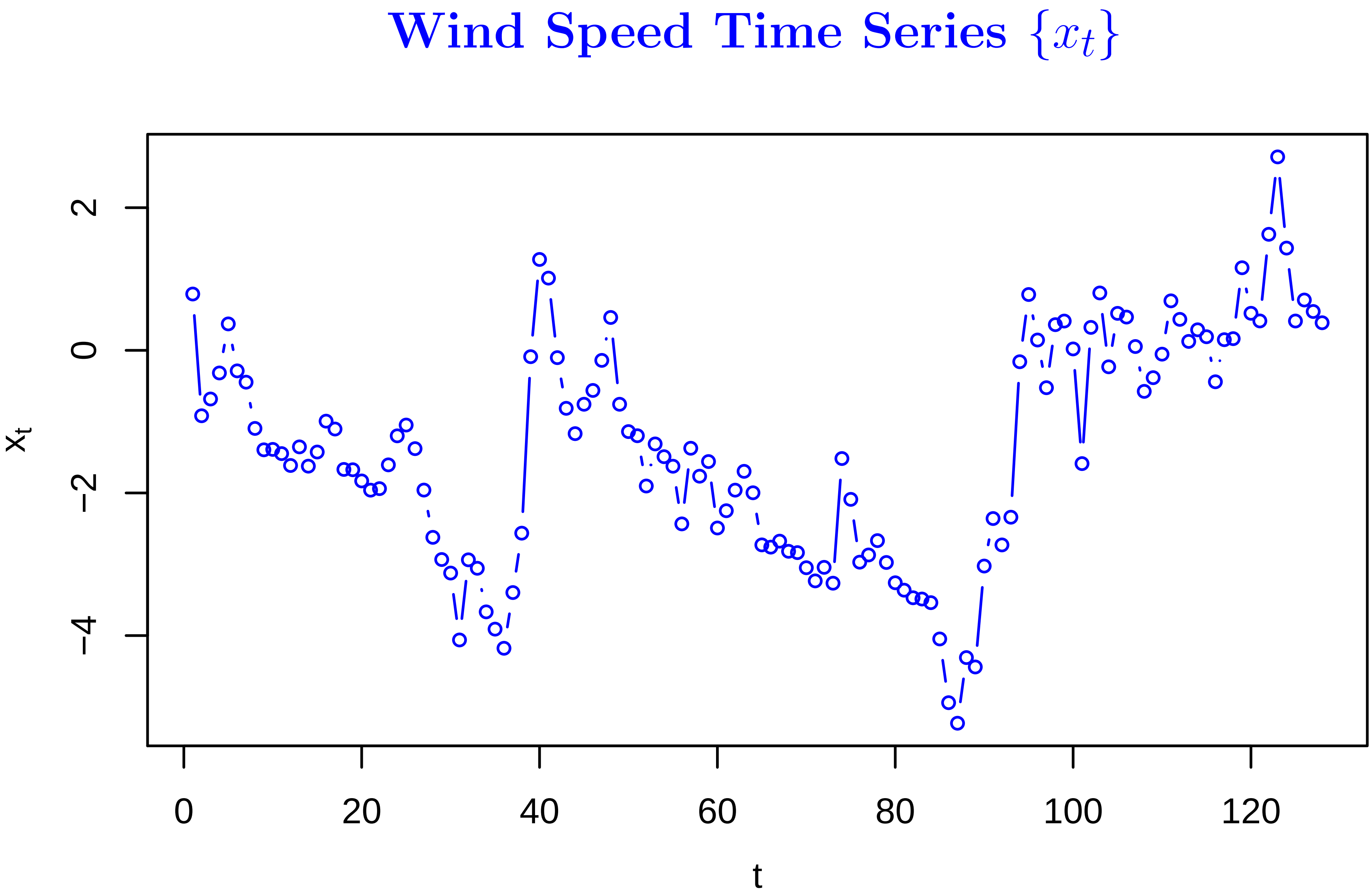
One-Step Ahead Prediction of AR(1) Process: II

- ϕX_n makes intuitive sense as a predictor since
$$X_{n+1} = \phi X_n + Z_{n+1} \text{ and } E\{Z_{n+1}\} = 0$$
- prediction error is $X_{n+1} - \phi X_n = Z_{n+1}$ and
$$\text{cov}\,\{Z_{n+1}, X_{n-j+1}\} = 0 \text{ for } j = 1, \ldots, n,$$
so orthogonality principle says ϕX_n is best linear predictor
- MSE is
$$\text{var}\,\{X_{n+1} - \widehat{X}_{n+1}\} = \gamma(0) - \boldsymbol{a}_n' \boldsymbol{\gamma}_n(1) = 1 - \phi^2$$
since $\boldsymbol{a}_n = [\phi, 0, \ldots, 0]'$ & $\boldsymbol{\gamma}_n(1) = [\gamma(1), \ldots, \gamma(n)]' = [\phi, \ldots, \phi^n]'$
- alternatively, noting that $X_{n+1} - \widehat{X}_{n+1} = X_{n+1} - \phi X_n = Z_{n+1}$, MSE is
$$\text{var}\,\{Z_{n+1}\} = 1 - \phi^2, \text{ as before}$$

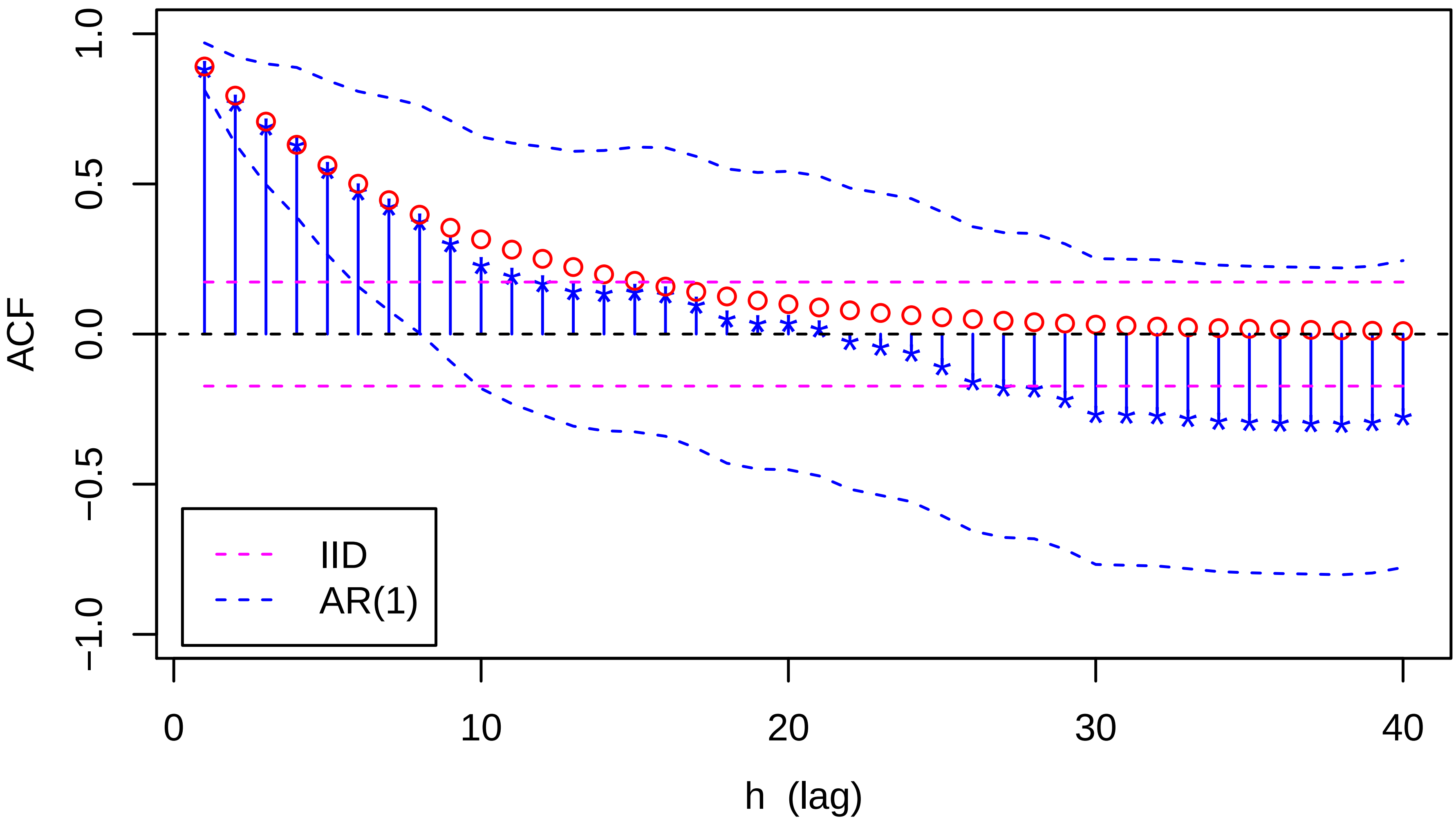
Example – Wind Speed Time Series

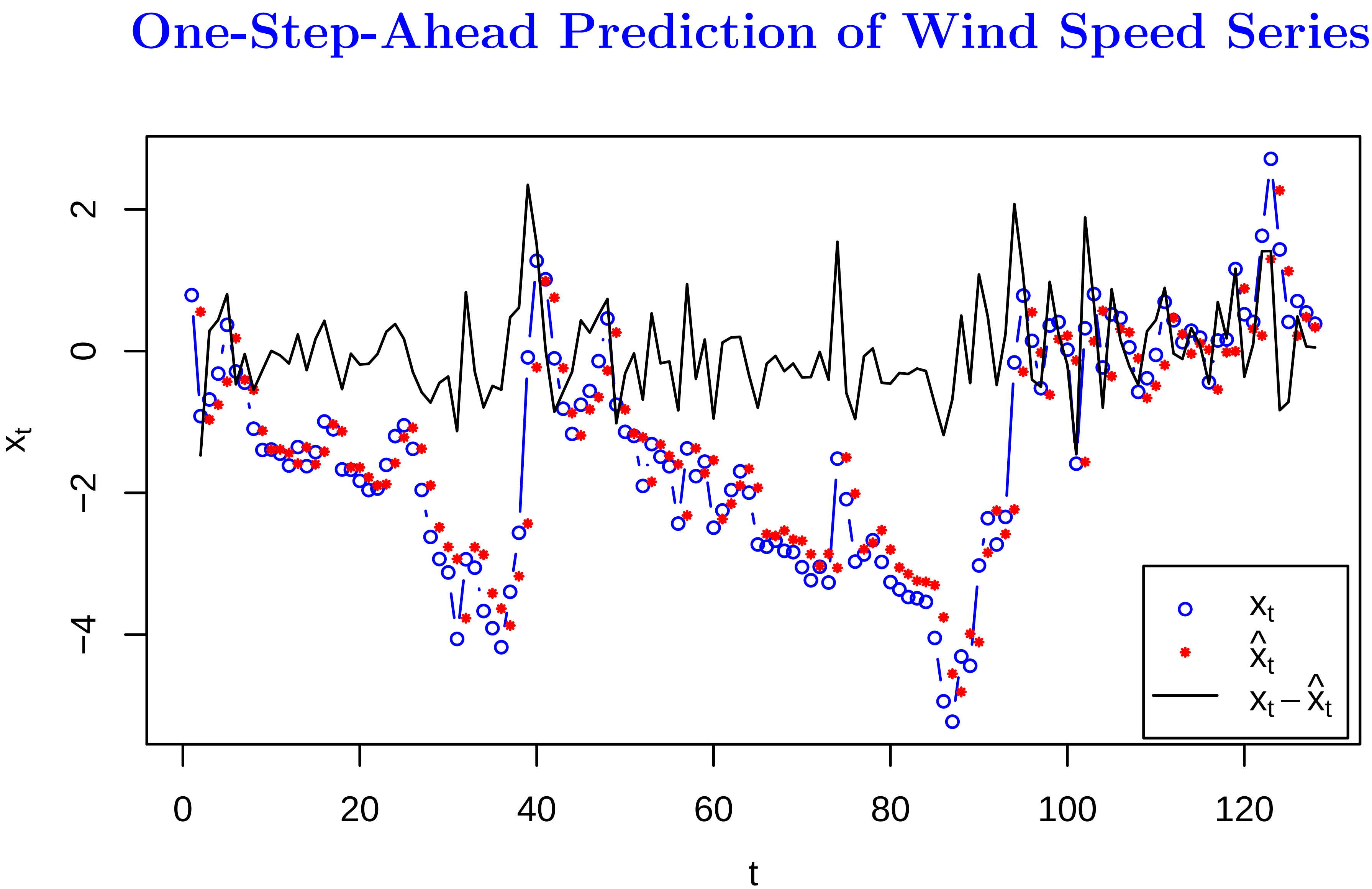
- sample ACF for wind speed series $\{x'_t\}$ indicates compatibility with AR(1) model, so let's use this series to illustrate forecasting one step ahead
- before doing so, need to take $\{x'_t\}$ and center it by subtracting off sample mean $\bar{x}' \doteq -1.37$ to form $\{x_t\}$, where $x_t = x'_t - \bar{x}$
- using $\hat{\phi} = \hat{\rho}(1) \doteq 0.891$, can predict x_t using $\hat{\phi} x_{t-1}$, and form observed prediction errors $z_t = x_t - \hat{\phi} x_{t-1}$ for $t = 2, 3, \ldots, n$
- estimated MSE and sample variance for $\{x_t\}$ are

$$\frac{1}{n-1} \sum_{t=2}^{n} z_t^2 \doteq 0.459 \text{ and } \frac{1}{n} \sum_{t=1}^{n} x_t^2 = \frac{1}{n} \sum_{t=1}^{n} (x'_t - \bar{x}')^2 \doteq 2.483$$


Wind Speed Time Series $\{x_t\}$
x_t
t
2
0
−2
−4
0
20
40
60
80
100
120

Model & Sample ACFs & 95% Confidence Bounds




One-Step-Ahead Prediction of Wind Speed Series
x_t
$\hat{x}_t$
$x_t - \hat{x}_t$
t
2
0
−2
−4
0
20
40
60
80
100
120

Generalization of Prediction Operation: I

- can easily generalize above results to handle case where we want to 'predict' RV Y based upon RVs $W_n, \ldots, W_1$ ('predict' in quotes because Y might occur temporally *before* some W_i's!)
- as before, can assume without loss of generality that RVs Y and W_i's all have mean zero (if not, just replace them with $Y - E\{Y\}$ and $W_i - E\{W_i\}$, which *do* have mean zero)
- some notation:
  - $\boldsymbol{W} = [W_1, W_2, \ldots, W_n]'$
  - $\boldsymbol{\gamma} = [\text{cov}\,\{Y, W_1\}, \text{cov}\,\{Y, W_2\}, \ldots, \text{cov}\,\{Y, W_n\}]'$
  - Γ is $n \times n$ covariance matrix for $\boldsymbol{W}$, i.e., its (i, j)th element is $\text{cov}\,\{W_i, W_j\}$
  - $\widehat{Y}$ is best linear predictor of Y given $\boldsymbol{W}$

Generalization of Prediction Operation: II

- using same arguments as before, have

$$\widehat{Y} = \boldsymbol{a}'\boldsymbol{W},$$

  where $\boldsymbol{a} = [a_1, a_2, \ldots, a_n]'$ is a solution to $\Gamma\boldsymbol{a} = \boldsymbol{\gamma}$
- corresponding MSE of prediction is

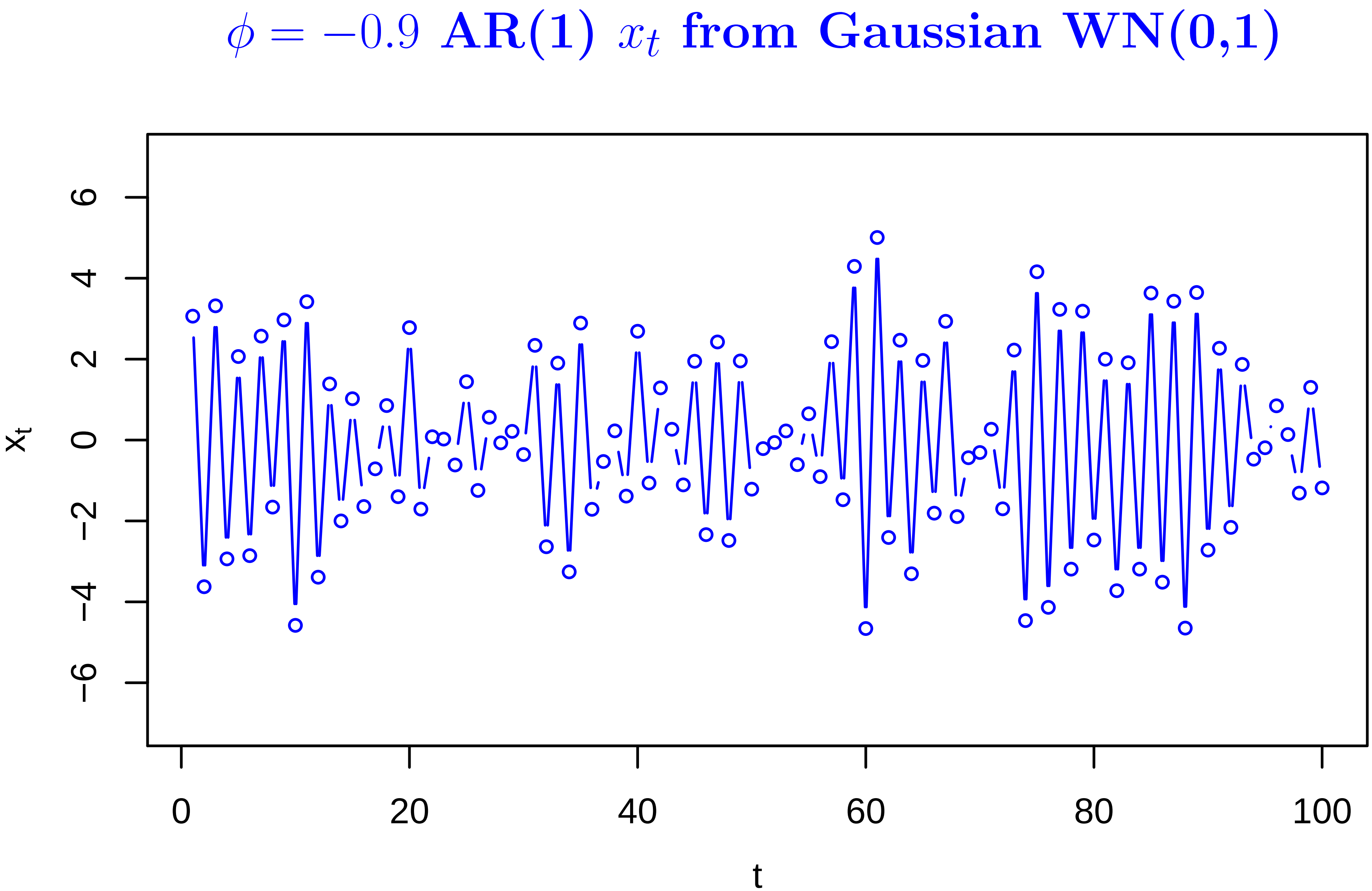
$$E\{(Y - \widehat{Y})^2\} = \operatorname{var}\{Y\} - \boldsymbol{a}'\boldsymbol{\gamma}$$

- as an example, suppose $\{X_t\}$ is a zero-mean stationary process with ACVF $\{\gamma_X(h)\}$ and ACF $\{\rho_X(h)\}$
- suppose we want to 'predict' $Y = X_2$ based on $W_1 = X_1$ and $W_2 = X_3$ (i.e., fill in a missing value)

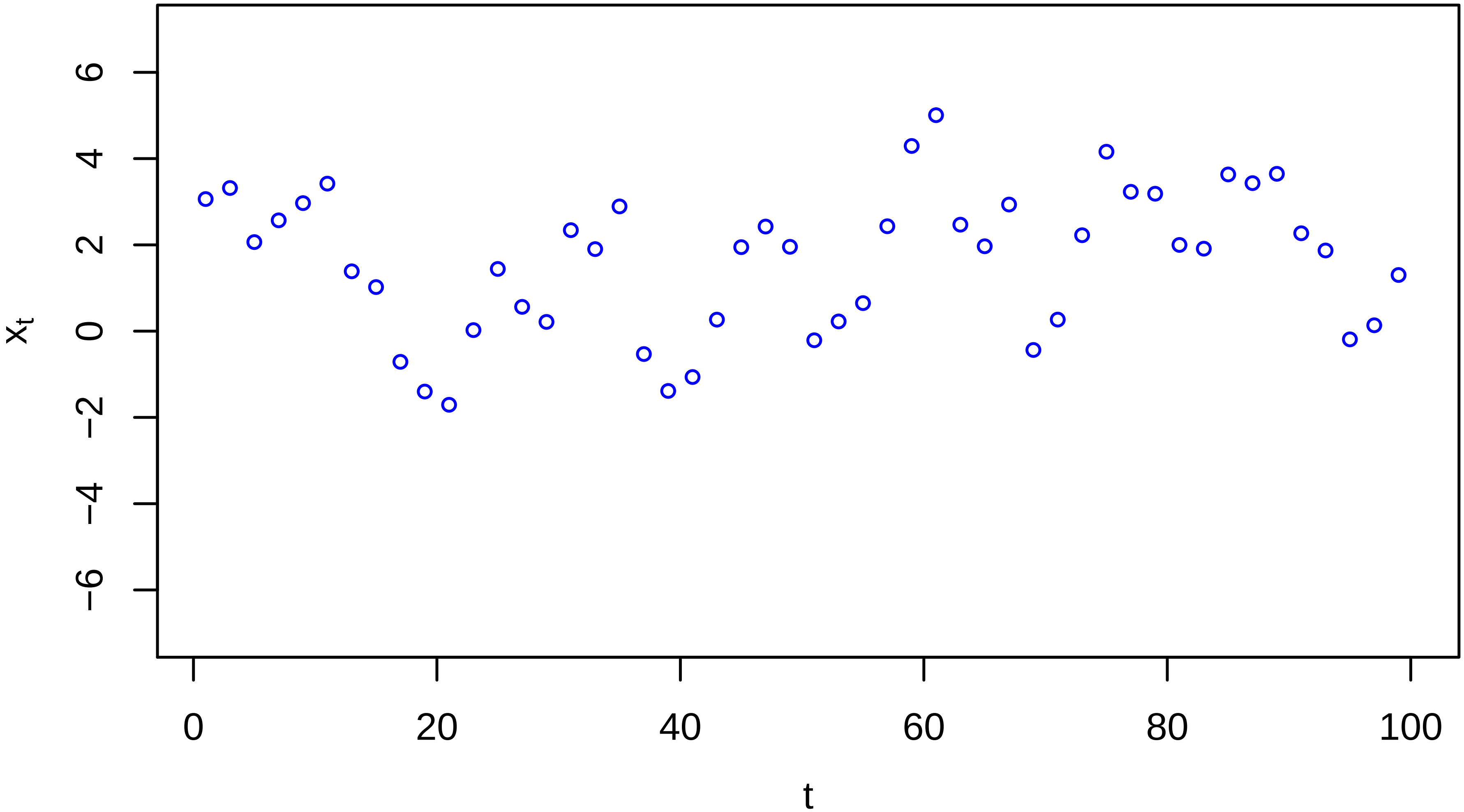
Generalization of Prediction Operation: III

- exercise: best linear predictor is
$$\widehat{X}_2 = \frac{\rho_X(1)}{1+\rho_X(2)}(X_1 + X_3)$$
with MSE
$$E\{(X_2 - \widehat{X}_2)^2\} = \gamma_X(0)\left(1 - \frac{2\rho_X^2(1)}{1+\rho_X(2)}\right)$$
(note: case $\rho_X(2) = -1$ requires special attention)

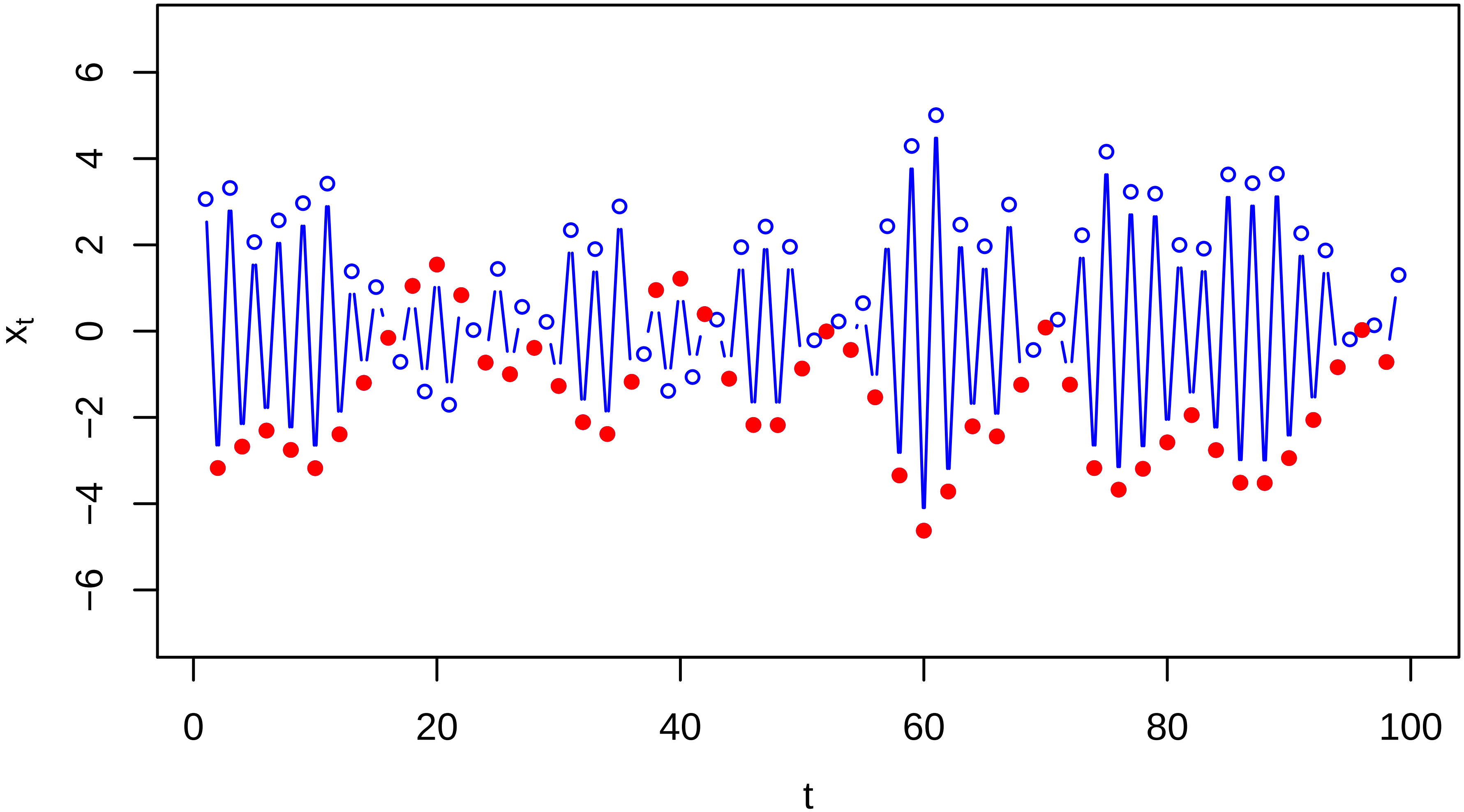
- specialize now to AR(1) process $X_t = \phi X_{t-1} + Z_t$, where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, thus yielding $\rho_X(h) = \phi^{|h|}$ and $\gamma(0) = \sigma^2/(1-\phi^2)$ and hence
$$\widehat{X}_2 = \frac{\phi}{1+\phi^2}(X_1 + X_3) \quad \text{with MSE} = \frac{\sigma^2}{1+\phi^2}$$

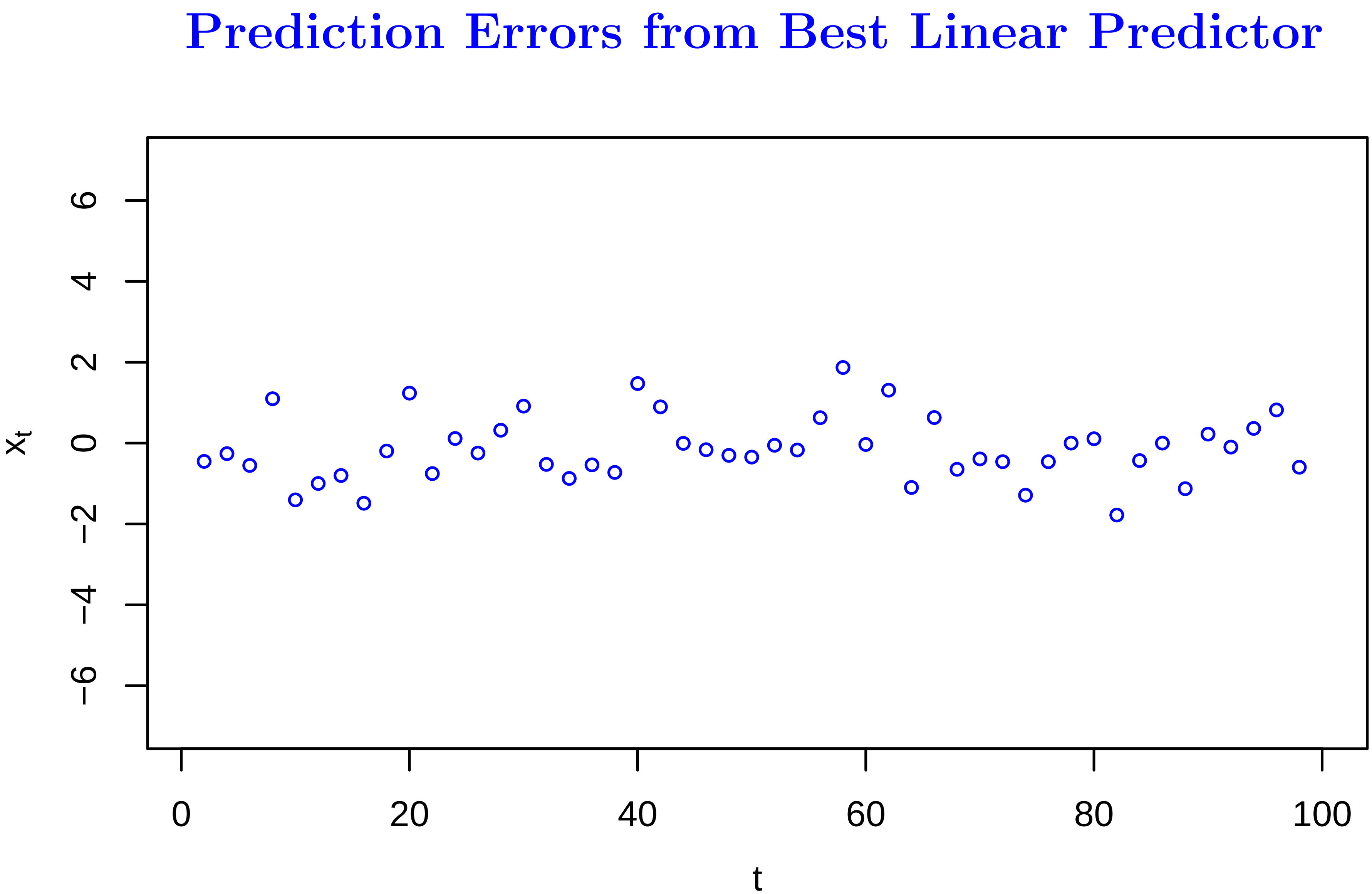

$\phi = -0.9$ AR(1) x_t from Gaussian WN(0,1)
6
4
2
0
−2
−4
−6
x_t
0
20
40
60
80
100
t

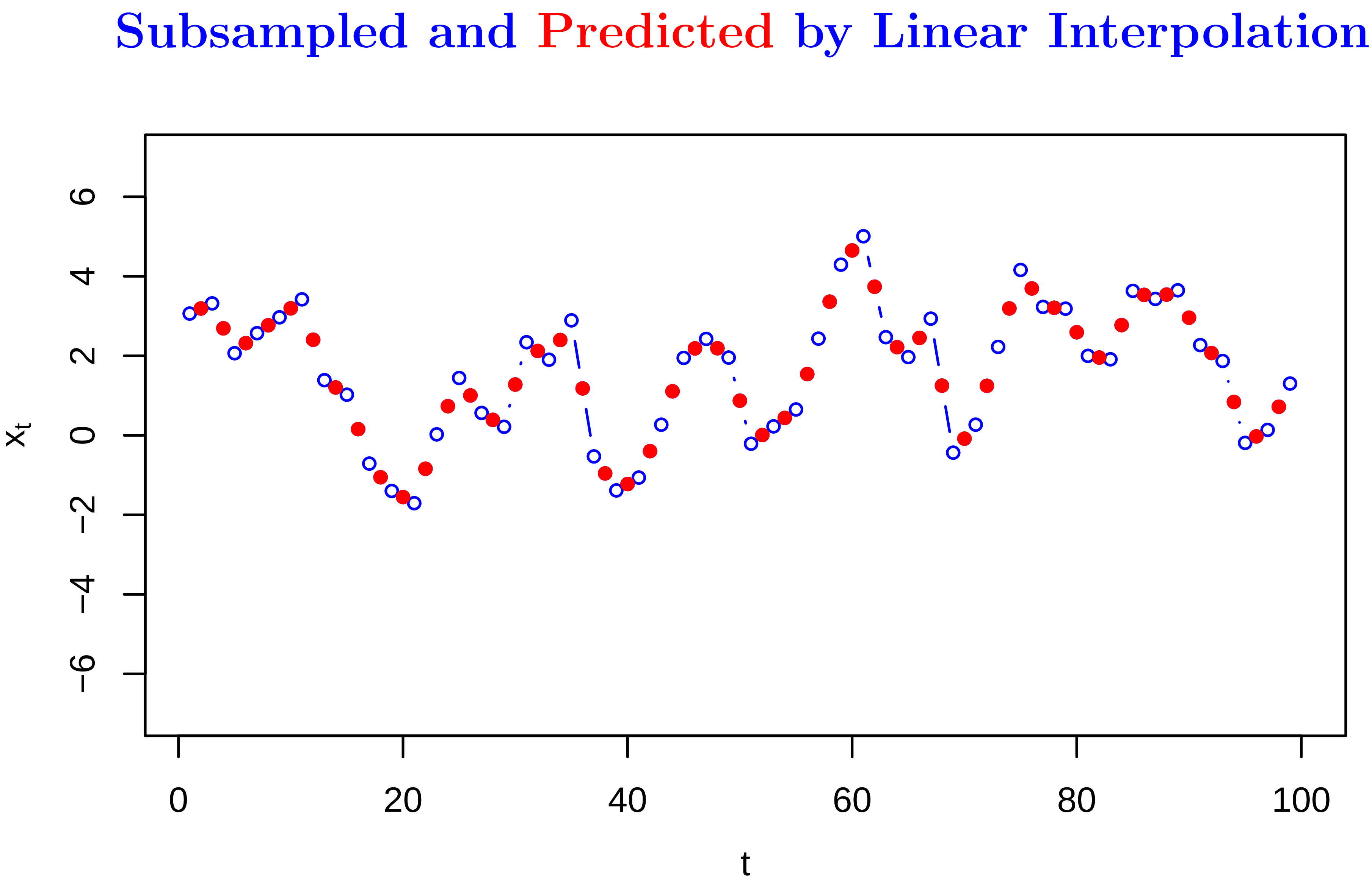
Subsampled $\phi = -0.9$ AR(1) $x_1, x_3, \ldots$



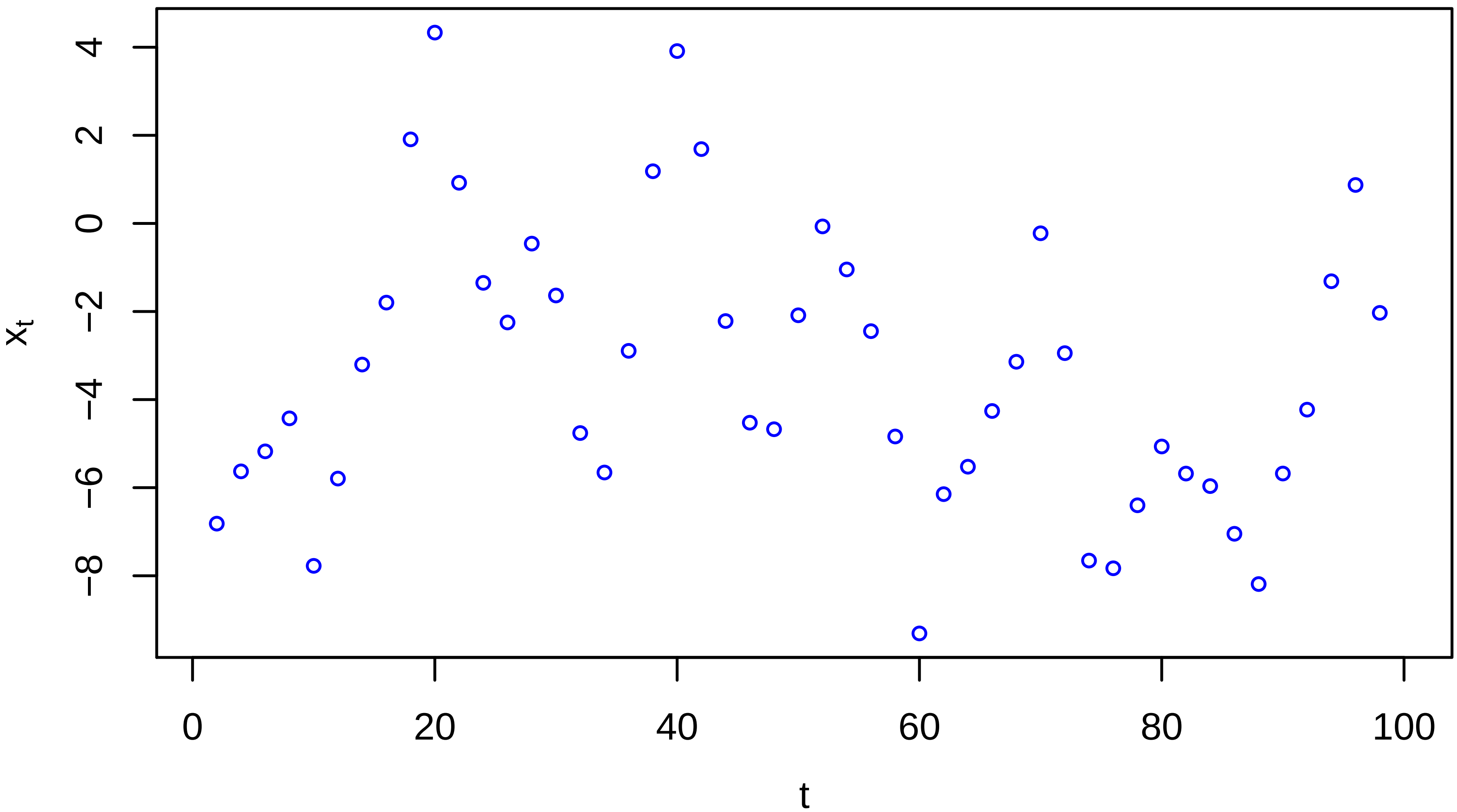
Subsampled and Predicted $\phi = -0.9$ AR(1) $x_1, x_3, \ldots$

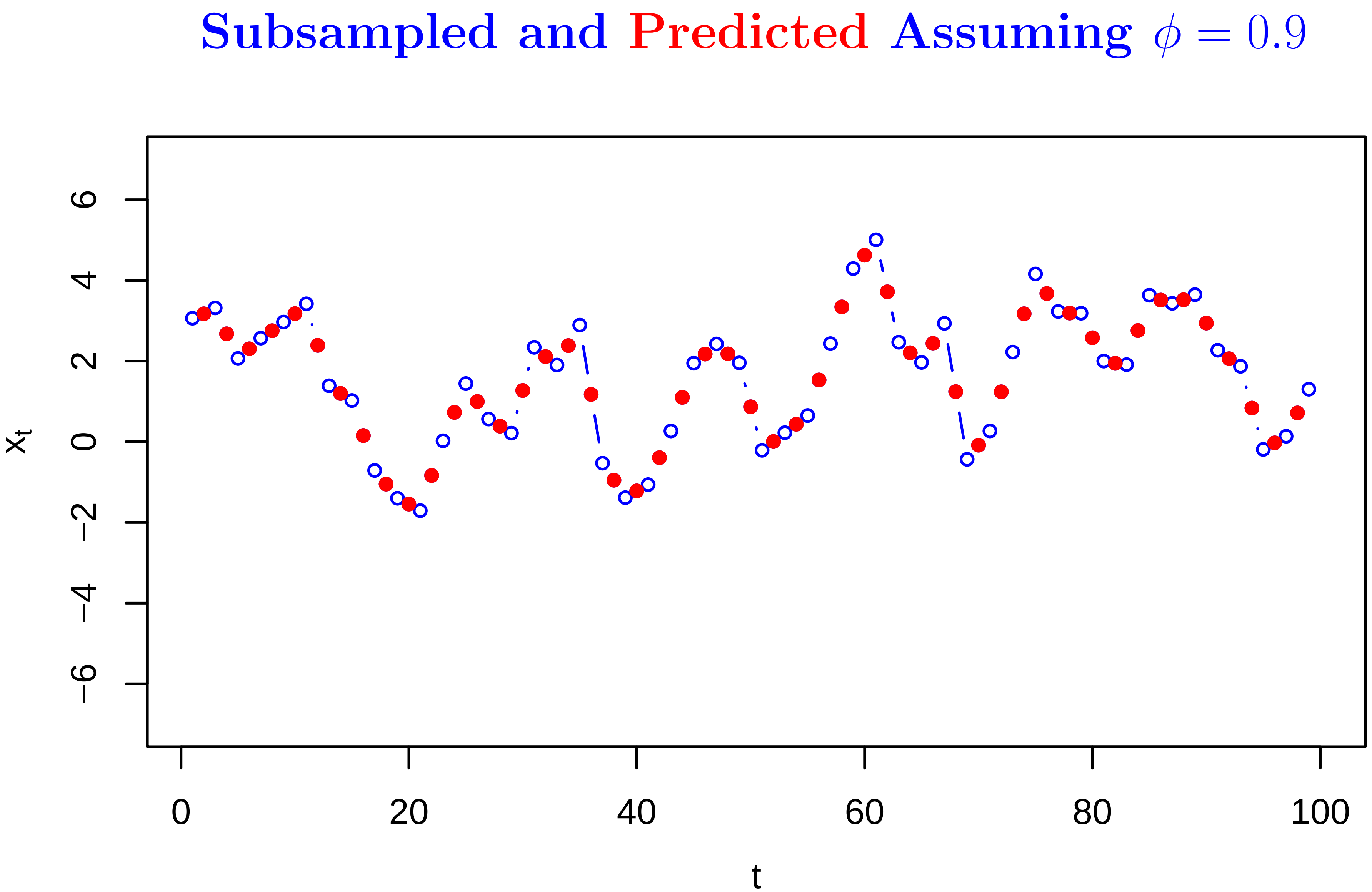



Prediction Errors from Best Linear Predictor
6
4
2
0
−2
−4
−6
x_t
0
20
40
60
80
100
t


Subsampled and Predicted by Linear Interpolation
x_t
6
4
2
0
−2
−4
−6
0
20
40
60
80
100
t

Prediction Errors from Linear Interpolation




Subsampled and Predicted Assuming $\phi = 0.9$
x_t
6
4
2
0
−2
−4
−6
0
20
40
60
80
100
t