Levinson–Durbin Recursions: I

- note: B&D and S&S say 'Durbin–Levinson' but 'Levinson–Durbin' is more commonly used (Levinson, 1947, and Durbin, 1960, are source articles – sometimes just 'Levinson' is used)
- recursions solve $\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(1)$ efficiently, giving us the coefficients $\boldsymbol{a}_n$ for best linear predictor $\widehat{X}_{n+1} = \boldsymbol{a}_n' \boldsymbol{X}_n$ of X_{n+1} given $\boldsymbol{X}_n = [X_n, \ldots, X_1]'$ under assumption that $E\{X_t\} = 0$
- in doing so, L–D recursions also give us
  - coefficients $\boldsymbol{a}_m$ for $\widehat{X}_{m+1} = \boldsymbol{a}_m' \boldsymbol{X}_m$, $m = 1, \ldots, n-1$, the best linear predictor of X_{m+1} given $\boldsymbol{X}_m = [X_m, \ldots, X_1]'$
  - partial autocorrelation function (PACF), also known as partial autocorrelation sequence or reflection coefficient sequence
- will state L–D recursions without proof (B&D have one; S&S leave it as exercise; Papoulis (1985) has an interesting one)

Levinson–Durbin Recursions: II

- to keep track of best linear predictors as sample size n increases (and to emphasize certain connections with AR processes), will switch notation from a_i to $\phi_{n,i}$
- henceforth we now write

$$\widehat{X}_{n+1} = \phi_{n,1}X_n + \phi_{n,2}X_{n-1} + \cdots + \phi_{n,n}X_1 = \boldsymbol{\phi}_n' \boldsymbol{X}_n$$

where $\boldsymbol{\phi}_n \stackrel{\text{def}}{=} [\phi_{n,1}, \phi_{n,1}, \ldots, \phi_{n,n}]'$

- simplify $\boldsymbol{\gamma}_n(1)$ to just $\boldsymbol{\gamma}_n$ so that $\boldsymbol{\gamma}_n = [\gamma(1), \gamma(2), \ldots, \gamma(n)]'$
- in new notation, L–D recursions solve for $\boldsymbol{\phi}_n$ in $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$
  - recall that Γ_n is covariance matrix for $\boldsymbol{X}_n$, so its (i, j)th element is $\text{cov}\,\{X_i, X_j\} = \gamma(i - j)$

Levinson–Durbin Recursions: III

- referring back to overheads X–11 to X–13, will denote mean square error (MSE) associated with predictor $\widehat{X}_{n+1}$ as

$$
\begin{aligned}
v_n &\stackrel{\text{def}}{=} E\{(X_{n+1} - \widehat{X}_{n+1})^2\} \\
&= \operatorname{var}\{X_{n+1} - \widehat{X}_{n+1}\} \\
&\stackrel{3.}{=} \gamma(0) - \boldsymbol{\phi}_n' \boldsymbol{\gamma}_n \\
&= \operatorname{var}\{X_{n+1}\} - \boldsymbol{\phi}_n' \operatorname{cov}\{X_{n+1}, \boldsymbol{X}_n\}
\end{aligned}
$$

Levinson–Durbin Recursions: IV

- for $n = 1$, have $\widehat{X}_2 \stackrel{\text{def}}{=} \phi_{1,1} X_1$
- equation $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$ becomes $\gamma(0)\phi_{1,1} = \gamma(1)$ $\quad (*)$
- solution is $\phi_{1,1} = \gamma(1)/\gamma(0) = \rho(1)$
- associated MSE is

$$\begin{aligned} v_1 &= \gamma(0) - \boldsymbol{\phi}_1' \boldsymbol{\gamma}_1 \\ &= \gamma(0) - \phi_{1,1}\gamma(1) \\ &= \gamma(0) - \phi_{1,1}[\phi_{1,1}\gamma(0)] \quad \text{(making use of } (*)) \\ &= \gamma(0)(1 - \phi_{1,1}^2) = v_0(1 - \phi_{1,1}^2) \text{ with } v_0 \stackrel{\text{def}}{=} \gamma(0) \end{aligned}$$

- Q: why is $\gamma(0)$ a natural definition for v_0?
- note connection to AR(1) model $X_t = \phi_{1,1} X_{t-1} + Z_t$ with $\{Z_t\} \sim \text{WN}(0, \sigma_X^2(1 - \phi_{1,1}^2))$, for which $\gamma(0) = \sigma_X^2$

Levinson–Durbin Recursions: V

- given $\boldsymbol{\phi}_{n-1}$ & v_{n-1}, L–D recursion gets $\boldsymbol{\phi}_n$ & v_n in 3 steps
  1. get *nth order partial autocorrelation* (more on this later!):
  $$\phi_{n,n} = \frac{\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)}{v_{n-1}}$$
  note: sum is inner product of $\boldsymbol{\phi}_{n-1}$ & order reversal of $\boldsymbol{\gamma}_{n-1}$
  2. get remaining $\phi_{n,j}$'s:
  $$\begin{bmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{n,n} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$
  3. get nth order MSE:
  $$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

Levinson–Durbin Recursions: VI

- as a first example, reconsider AR(1) process $X_t = \phi X_{t-1} + Z_t$, where $|\phi| < 1$ and $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$
- have already argued (X–14) that $\widehat{X}_{n+1} = \phi X_n$
- $\boldsymbol{\phi}_n = [\phi, 0, \ldots, 0]'$ and $v_n = \sigma^2$ for all $n \geq 1$ since MSE is σ^2
- assuming $n \geq 2$, let's apply L–D recursions to $\boldsymbol{\phi}_{n-1} = [\phi, 0, \ldots, 0]'$ & $v_{n-1} = \sigma^2$ and see if required forms for $\boldsymbol{\phi}_n$ and v_n pop out
- step 1: recalling that $\gamma(h) = \sigma^2 \phi^h / (1 - \phi^2)$ for $h \geq 0$, we have

$$\begin{aligned}
\phi_{n,n} &= \frac{\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)}{v_{n-1}} \\
&= \sigma^2 \frac{\phi^n - \phi_{n-1,1}\phi^{n-1}}{v_{n-1}(1-\phi^2)} = \sigma^2 \frac{\phi^n - \phi^n}{v_{n-1}(1-\phi^2)} = 0
\end{aligned}$$

Levinson–Durbin Recursions: VII

- step 2:

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

yields

$$\begin{bmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ \phi_{n,n-2} \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} - 0 \times \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \phi \end{bmatrix} = \begin{bmatrix} \phi \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},$$

so $\boldsymbol{\phi}_n = [\phi, 0, \ldots, 0]'$ as required

Levinson–Durbin Recursions: VIII

- step 3: $v_n = v_{n-1}(1 - \phi_{n,n}^2) = v_{n-1} = \sigma^2$, as required
- note: partial autocorrelation $\phi_{n,n}$ for AR(1) process is ϕ for $n = 1$ and is zero for $n = 2, 3, \ldots$
- exercise: run L–D recursions on MA(1) process
- as 2nd example, reconsider stationary process of Problem 2(b):

$$X_t = Z_2 \cos(\omega t) + Z_1 \sin(\omega t),$$

  where Z_1 and Z_2 are independent $\mathcal{N}(0, 1)$ RVs
- ACVF for $\{X_t\}$ is $\gamma(h) = \cos(\omega h)$ (same as is its ACF $\rho(h)$)
- starting with $\widehat{X}_2 \stackrel{\text{def}}{=} \phi_{1,1} X_1$ ($n = 1$ case), we have

$$\phi_{1,1} = \rho(1) = \cos(\omega) \text{ and } v_1 = \gamma(0)(1 - \phi_{1,1}^2) = 1 - \cos^2(\omega) = \sin^2(\omega)$$

Levinson–Durbin Recursions: IX

- now let us get coefficients for $\widehat{X}_3 \stackrel{\text{def}}{=} \phi_{2,1}X_2 + \phi_{2,2}X_1$ ($n = 2$ case) using L–D recursions
- first step

$$\phi_{n,n} = \frac{\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)}{v_{n-1}},$$

yields, for $n = 2$ (recalling $\gamma(h) = \cos(\omega h)$ & $\phi_{1,1} = \cos(\omega)$),

$$\phi_{2,2} = \frac{\gamma(2) - \phi_{1,1}\gamma(1)}{v_1} = \frac{\cos(2\omega) - \cos(\omega)\cos(\omega)}{\sin^2(\omega)} = -1$$

because of trig identity $\cos(2\omega) - \cos^2(\omega) = -\sin^2(\omega)$

Levinson–Durbin Recursions: X

- second step of L–D recursions, namely,
$$\begin{bmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{n,n} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix},$$
yields, for $n = 2$,
$$\phi_{2,1} = \phi_{1,1} - \phi_{2,2}\phi_{1,1} = \cos(\omega)[1 - (-1)] = 2\cos(\omega)$$
- third step of L–D recursions, namely,
$v_n = v_{n-1}(1 - \phi_{n,n}^2)$ yields, for $n = 2$, $v_2 = v_1[1 - (-1)^2] = 0$
- thus X_3 is perfectly predicable given X_2 & X_1:
$$\widehat{X}_3 = 2\cos(\omega)X_2 - X_1 = X_3$$
- thus, for all t, X_t is perfectly predicable given X_{t-1} & X_{t-2}:
$$\widehat{X}_t = 2\ \cos(\omega)X_{t-1} - X_{t-2} = X_t \quad \text{(Q: why?)}$$

Aside – Step-Down Levinson–Durbin Recursions: I

- application of L–D recursions to AR(p) process
$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$$
yields, for $n \geq p$,
$\widehat{Y}_{n+1} = \phi_{n,1} Y_n + \cdots + \phi_{n,n} Y_1 = \phi_1 Y_n + \cdots + \phi_p Y_{n-p+1}$,
i.e., $\widehat{Y}_{n+1}$ only depends on p most recent values and, when $n > p$, not on remote values $Y_{n-p}, \ldots, Y_1$
- associated prediction error is
$$Y_{n+1} - \widehat{Y}_{n+1} = Y_{n+1} - \phi_1 Y_n - \cdots - \phi_p Y_{n-p+1} = Z_{n+1},$$
so MSE is $v_n = \text{var}\{Y_{n+1} - \widehat{Y}_{n+1}\} = \text{var}\{Z_{n+1}\} = \sigma^2$
- given $\phi_{p,1} = \phi_1, \phi_{p,2} = \phi_2, \ldots, \phi_{p,p} = \phi_p$ and σ^2, can 'invert' L–D recursions to get coefficients for best linear predictors of orders $p-1, p-2, \ldots, 1$ and associated MSEs

Aside – Step-Down Levinson–Durbin Recursions: II

- given $\phi_{h,1}, \ldots, \phi_{h,h}$ & v_h, compute
  1. $\phi_{h-1,j} = \frac{\phi_{h,j}+\phi_{h,h}\phi_{h,h-j}}{1-\phi_{h,h}^2}$, $1 \le j \le h-1$
  2. $v_{h-1} = v_h/(1-\phi_{h,h}^2)$
- 'step-down' L–D recursion yields $\phi_{h-1,1}, \ldots, \phi_{h-1,h-1}$ & v_{h-1}
- start with $\phi_{p,1} = \phi_1, \ldots, \phi_{p,p} = \phi_p$ & $v_p = \sigma^2$
- apply step-down recursions to get

$$\phi_{p-1,j}\text{'s} \ \& \ v_{p-1}, \ \phi_{p-2,j}\text{'s} \ \& \ v_{p-2}, \ldots, \phi_{1,1} \ \& \ v_1$$

- as opposed to usual L–D recursions, step-down L–D recursions do *not* make use of ACVF $\gamma(h)$ for $\{Y_t\}$
- in fact, given $\phi_1, \phi_2, \ldots, \phi_p$ & σ^2, can use results of step-down L–D recursions to *compute* $\gamma(h)$ (yet another method!)

Aside – Step-Down Levinson–Durbin Recursions: III

- to do so, return to overhead XI–4 and note that

$$\begin{aligned}\gamma(0) &= v_0 = v_1/(1-\phi_{1,1}^2)\\ \gamma(1) &= \gamma(0)\phi_{1,1}\end{aligned}$$

- next go to overhead XI–5, grab

$$\phi_{n,n} = \frac{\gamma(n) - \sum_{j=1}^{n-1}\phi_{n-1,j}\gamma(n-j)}{v_{n-1}}$$

and manipulate it to get

$$\gamma(n) = \phi_{n,n}v_{n-1} + \sum_{j=1}^{n-1}\phi_{n-1,j}\gamma(n-j)$$

and thus $\gamma(2) = \phi_{2,2}v_1 + \phi_{1,1}\gamma(1)$

$\gamma(3) = \phi_{3,3}v_2 + \phi_{2,1}\gamma(2) + \phi_{2,2}\gamma(1)$ etc., ending with

$$\gamma(p) = \phi_{p,p}v_{p-1} + \phi_{p-1,1}\gamma(p-1) + \cdots + \phi_{p-1,p-1}\gamma(1)$$

Aside – Step-Down Levinson–Durbin Recursions: IV

- to get $\gamma(p+1), \gamma(p+2), \ldots$, make use of an equation stated on overhead IX–50:

$$\gamma(k) = \phi_1 \gamma(k-1) + \cdots + \phi_p \gamma(k-p),$$

  which holds for all $k \geq p+1$

- note: can now argue that AR coefficients

$$\phi_1, \phi_2, \ldots, \phi_p$$

  and sequence of partial autocorrelations

$$\phi_{1,1}, \phi_{2,2}, \ldots, \phi_{p,p}$$

  are equivalent to one another (in particular, $\phi_{p,p} = \phi_p$)

- we now return to our regularly scheduled program ...

One-Step-Ahead Prediction Errors (Innovations): I

- given time series X_1, X_2, ..., can use L–D recursions to find coefficients $\boldsymbol{\phi}_{m-1}$ for $\widehat{X}_m$ i.e., best linear predictor of X_m given $X_{m-1}, \ldots, X_1$
- define $\widehat{X}_1 = 0$ and $\widehat{\boldsymbol{X}}_n = [\widehat{X}_n, \widehat{X}_{n-1}, \ldots, \widehat{X}_1]'$
- letting $m = 1, 2, \ldots, n$, can generate a series of one-step-ahead prediction errors (or *innovations*):

$$U_m = X_m - \widehat{X}_m$$

- collect these into $\boldsymbol{U}_n = [U_n, U_{n-1} \ldots, U_1]'$ so that we can write

$$\boldsymbol{U}_n = \boldsymbol{X}_n - \widehat{\boldsymbol{X}}_n$$

One-Step-Ahead Prediction Errors (Innovations): II

- can write $\boldsymbol{U}_n = A_n'\boldsymbol{X}_n$, where A_n is lower triangular:

$$A_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\phi_{n-1,1} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ -\phi_{n-1,n-3} & -\phi_{n-2,n-4} & \cdots & 1 & 0 & 0 \\ -\phi_{n-1,n-2} & -\phi_{n-2,n-3} & \cdots & -\phi_{2,1} & 1 & 0 \\ -\phi_{n-1,n-1} & -\phi_{n-2,n-2} & \cdots & -\phi_{2,2} & -\phi_{1,1} & 1 \end{bmatrix}$$

- inverse of A_n is also lower triangular, so let's write it as

$$C_n \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \theta_{n-1,1} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ \theta_{n-1,n-3} & \theta_{n-2,n-4} & \cdots & 1 & 0 & 0 \\ \theta_{n-1,n-2} & \theta_{n-2,n-3} & \cdots & \theta_{2,1} & 1 & 0 \\ \theta_{n-1,n-1} & \theta_{n-2,n-2} & \cdots & \theta_{2,2} & \theta_{1,1} & 1 \end{bmatrix}$$

One-Step-Ahead Prediction Errors (Innovations): III

- since C_n' is inverse of A_n', $\boldsymbol{U}_n = A_n'\boldsymbol{X}_n$ leads to $\boldsymbol{X}_n = C_n'\boldsymbol{U}_n$; i.e., time series can be reexpressed in terms of its innovations
- recall that L–D recursions give

$$v_{m-1} = E\{(X_m - \widehat{X}_m)^2\} = \text{var}\{U_m\}, \quad m = 1, 2, \ldots, n$$

- can use *innovations algorithm* (IA) to get v_m and elements of C_m (note: take sum with upper limit '−1' to be 0):

$$\theta_{m,m-k} = \frac{\gamma(m-k) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{m,m-j}v_j}{v_k}, \quad 0 \leq k < m$$

$$v_m = \gamma(0) - \sum_{j=0}^{m-1} \theta_{m,m-j}^2 v_j$$

One-Step-Ahead Prediction Errors (Innovations): IV

- starting with $v_0 = \gamma(0)$, here are $m = 1$ and 2 steps of IA:

$$
\begin{aligned}
\theta_{1,1} &= \frac{\gamma(1)}{v_0} \\
v_1 &= \gamma(0) - \theta_{1,1}^2 v_0 \\
\theta_{2,2} &= \frac{\gamma(2)}{v_0} \\
\theta_{2,1} &= \frac{\gamma(1) - \theta_{1,1}\theta_{2,2}v_0}{v_1} \\
v_2 &= \gamma(0) - \theta_{2,2}^2 v_0 - \theta_{2,1}^2 v_1
\end{aligned}
$$

One-Step-Ahead Prediction Errors (Innovations): V

- since $\boldsymbol{X}_n = C_n' \boldsymbol{U}_n$, can write (with $\theta_{m,0} \stackrel{\text{def}}{=} 1$ for $m \geq 0$),

$$X_{m+1} = \sum_{j=0}^{m} \theta_{m,j} U_{m-j+1}, \quad m = 0, 1, \ldots, n-1,$$

  i.e., linear combination of innovations yields time series

- since $\widehat{\boldsymbol{X}}_n = \boldsymbol{X}_n - \boldsymbol{U}_n = C_n' \boldsymbol{U}_n - \boldsymbol{U}_n = (C_n' - I_n)\boldsymbol{U}_n$, where I_n is the $n \times n$ identity matrix, have $\widehat{X}_0 = 0$ (its definition) and

$$\widehat{X}_{m+1} = \sum_{j=1}^{m} \theta_{m,j} U_{m-j+1}, \quad m = 1, 2, \ldots, n-1,$$

  i.e., linear combination of innovations also yields predictions

- exercise: innovations $U_1, U_2, \ldots, U_n$ are uncorrelated

Aside – Simulation of ARMA Processes: I

- suppose we want to generate realizations of an ARMA process
- look first at causal (& thus stationary) Gaussian AR(p) process:

$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = Z_t, \ \{Z_t\} \sim$ Gaussian WN(0,σ^2)

- recall that, for any $t \geq p+1$, best linear predictor $\widehat{Y}_t$ of Y_t given $Y_{t-1}, \ldots, Y_1$ takes the form

$$\widehat{Y}_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}$$

- innovations are $U_t = Y_t - \widehat{Y}_t = Z_t$ and have MSE

$$v_{t-1} \stackrel{\text{def}}{=} \text{var}\,\{U_t\} = \sigma^2$$

- can use step-down L–D recursions to get coefficients for

$$\widehat{Y}_t = \phi_{t-1,1} Y_{t-1} + \cdots + \phi_{t-1,t-1} Y_1, \ \ t = 2, 3, \ldots, p$$

and associated MSEs v_{t-1} (recall that $\hat{Y}_1 = 0$ by definition)

Aside – Simulation of ARMA Processes: II

- innovations $U_t = Y_t - \widehat{Y}_t$, $t = 1, \ldots, p$, are such that
  1. $E\{U_t\} = 0$ and $\text{var}\,\{U_t\} = v_{t-1}$
  2. $U_1,\ U_2,\ \ldots,\ U_p$ are uncorrelated RVs (exercise) – implies independence under Gaussian assumption
- easy to simulate U_t's: generate p independent realizations of $\mathcal{N}(0,1)$ RVs, say, $\widetilde{Z}_1, \ldots, \widetilde{Z}_p$, and set $U_t = v_{t-1}^{1/2}\widetilde{Z}_t$
- can 'unroll' U_t's to get simulations of Y_t's, $t = 1, \ldots, p$:

$$U_1 = Y_1 - \widehat{Y}_1 = Y_1 \text{ yields } Y_1 = U_1$$
$$U_2 = Y_2 - \widehat{Y}_2 = Y_2 - \phi_{1,1}Y_1 \text{ yields } Y_2 = \phi_{1,1}Y_1 + U_2$$
$$U_3 = Y_3 - \widehat{Y}_3 = Y_3 - \phi_{2,1}Y_2 - \phi_{2,2}Y_1 \text{ yields } Y_3 = \phi_{2,1}Y_2 + \phi_{2,2}Y_1 + U_3$$

Aside – Simulation of ARMA Processes: III

- finally
$$U_p = Y_p - \widehat{Y}_p = Y_p - \phi_{p-1,1}Y_{p-1} - \cdots - \phi_{p-1,p-1}Y_1$$
yields
$$Y_p = \phi_{p-1,1}Y_{p-1} + \cdots + \phi_{p-1,p-1}Y_1 + U_p$$
- can now generate remainder of desired simulated series using
$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \sigma \widetilde{Z}_t, \ \ t = p+1, p+2, \ldots,$$
where $\widetilde{Z}_t$'s are independent realizations of $\mathcal{N}(0,1)$ RVs (these are independent of $\widetilde{Z}_1, \ldots, \widetilde{Z}_p$ also)

Aside – Simulation of ARMA Processes: IV

- knowing how to simulate AR process $\phi(B)Y_t = Z_t$, can in turn simulate ARMA process

$$\phi(B)X_t = \theta(B)Z_t$$

  since we can create ARMA process $\{X_t\}$ by applying filter $\theta(B)$ to AR process $\{Y_t\}$:

$$X_t = \theta(B)Y_t = \theta(B)\phi^{-1}(B)Z_t, \text{ i.e., } \phi(B)X_t = \theta(B)Z_t$$

  (see overhead IX–47)

- hence can generate simulated ARMA series of length n via

$$X_t = Y_t + \theta_1 Y_{t-1} + \cdots + \theta_q Y_{t-q}, \quad t = q+1, \ldots, q+n;$$

  i.e., need to make simulated AR series of length $n + q$

Example – Simulation of ARMA(2,2) Process: I

- consider ARMA(2,2) process given by

$$X_t = \tfrac{3}{4}X_{t-1} - \tfrac{1}{2}X_{t-2} + Z_t + \tfrac{7}{10}Z_{t-1} - \tfrac{1}{10}Z_{t-2}, \ \{Z_t\} \sim \text{WN}(0,1)$$

- to simulate AR(2) process

$$Y_t = \tfrac{3}{4}Y_{t-1} - \tfrac{1}{2}Y_{t-2} + Z_t \text{ for which } v_2 = 1,$$

need to run reverse L–D recursions once to obtain

$$\phi_{1,1} = \frac{\phi_{2,1} + \phi_{2,2}\phi_{2,1}}{1 - \phi_{2,2}^2} = \frac{\frac{3}{4} - \frac{1}{2} \times \frac{3}{4}}{1 - \frac{1}{4}} = \tfrac{1}{2},$$

$$v_1 = \frac{v_2}{1 - \phi_{2,2}^2} = \tfrac{4}{3} \text{ and hence } v_0 = \frac{v_1}{1 - \phi_{1,1}^2} = \tfrac{16}{9}$$

Example – Simulation of ARMA(2,2) Process: II

- thus would generate AR(2) process using

$$\begin{aligned} Y_1 &= \tfrac{4}{3}\widetilde{Z}_1 \\ Y_2 &= \tfrac{1}{2}Y_1 + \tfrac{2}{\sqrt{3}}\widetilde{Z}_2 \\ Y_3 &= \tfrac{3}{4}Y_2 - \tfrac{1}{2}Y_1 + \widetilde{Z}_3 \\ &\vdots \\ Y_{n+2} &= \tfrac{3}{4}Y_{n+1} - \tfrac{1}{2}Y_n + \widetilde{Z}_{n+2}, \end{aligned}$$

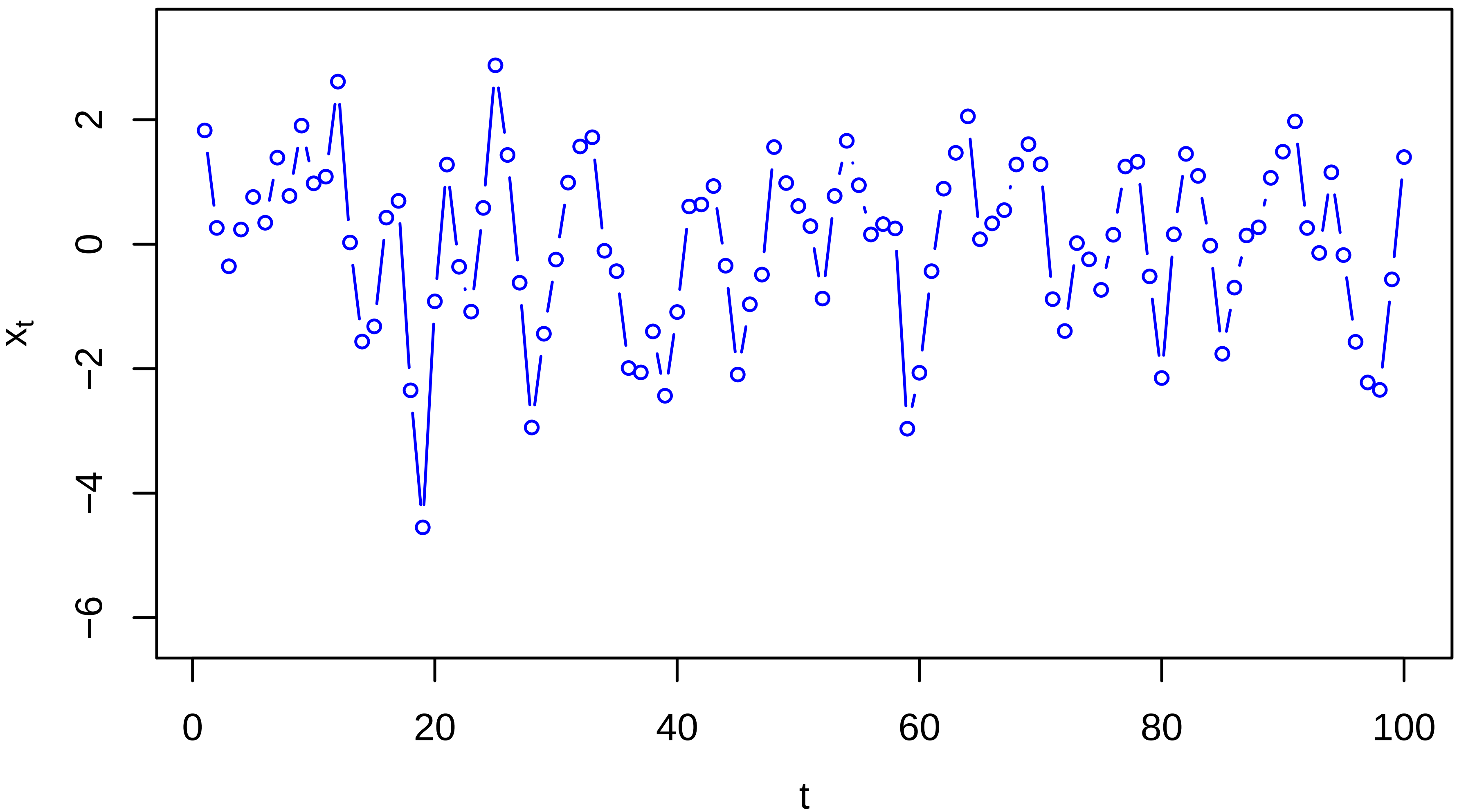
  where $\widetilde{Z}_t$'s are IID $\mathcal{N}(0,1)$ RVs

- desired ARMA(2,2) process is given by

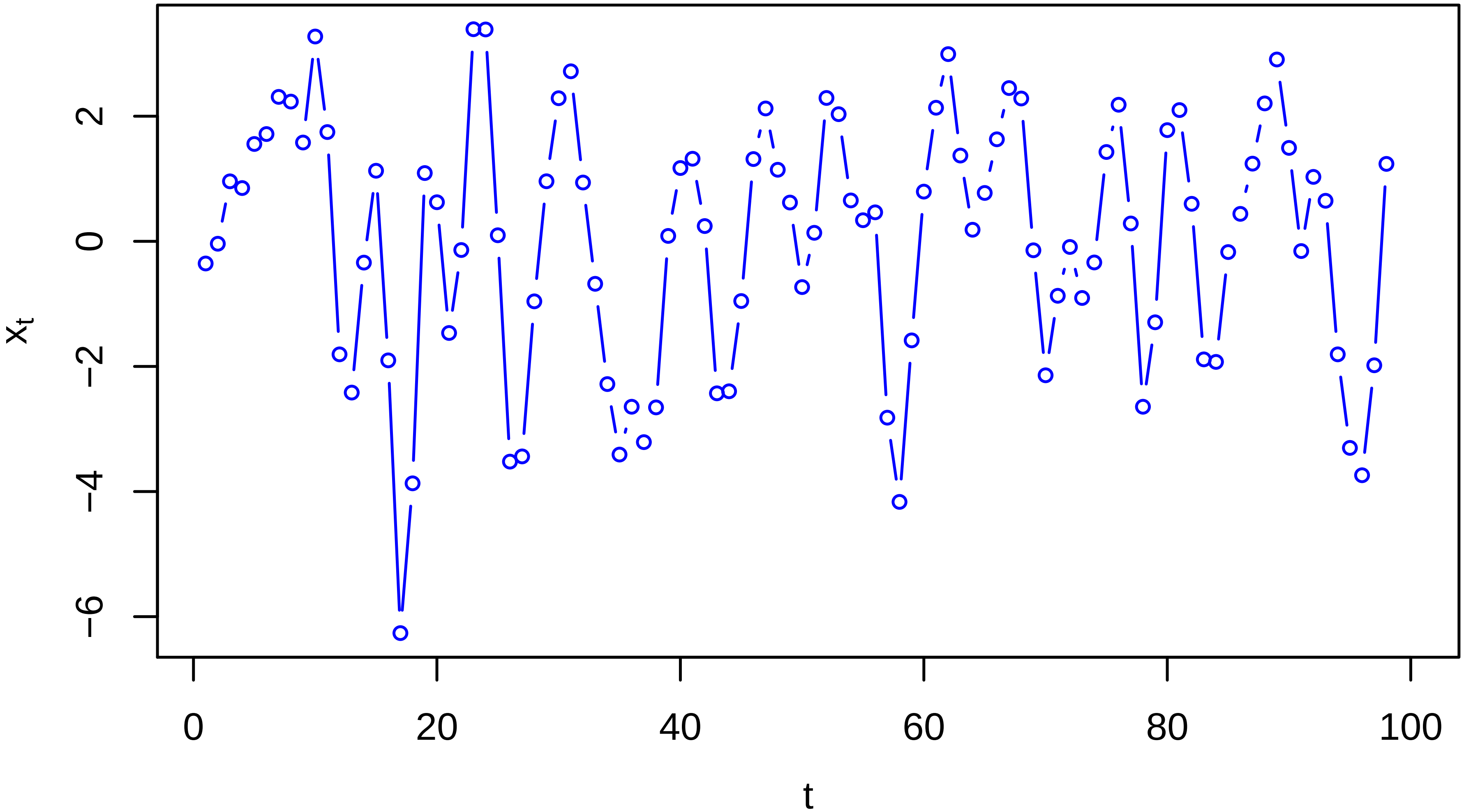
$$X_t = Y_{t+2} + \tfrac{7}{10}Y_{t+1} - \tfrac{1}{10}Y_t, \ \ t = 1, \ldots, n$$

- overhead VIII–28 shows AR(2) series ($n = 100$), which we can use to form a ARMA(2,2) simulation ($n = 98$)

Realization of Second AR(2) Process ($n = 100$)



Realization of ARMA(2,2) Process ($n = 98$)



Aside – Simulation of ARMA Processes: V

- method described here deemed 'exact' because of use of so-called stationary initial conditions (method used in `R` function `arima.sim` is not exact – makes use of a 'burn-in' period)
- source article is Kay (1981), which is just over a page in length, making it one of the shortest useful articles relevant to time series analysis (shortest is undoubtedly David, 1985!)

Aside – Simulation of ARMA Processes: VI

- in AR(p) case, can regard simulation procedure as based directly on $\boldsymbol{U}_n = A'_n \boldsymbol{X}_n$:
  - use L–D to get elements of A'_n & variances of innovations
  - simulate innovations $\boldsymbol{U}_n$ (easy to do!)
  - starting at bottom, unravel to get desired simulation of $\boldsymbol{X}_n$
- as example, consider AR(1) process $X_t = \phi X_{t-1} + Z_t$ & $n = 6$:

$$\begin{bmatrix} U_6 \\ U_5 \\ U_4 \\ U_3 \\ U_2 \\ U_1 \end{bmatrix} = \begin{bmatrix} 1 & -\phi & 0 & 0 & 0 & 0 \\ 0 & 1 & -\phi & 0 & 0 & 0 \\ 0 & 0 & 1 & -\phi & 0 & 0 \\ 0 & 0 & 0 & 1 & -\phi & 0 \\ 0 & 0 & 0 & 0 & 1 & -\phi \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_6 \\ X_5 \\ X_4 \\ X_3 \\ X_2 \\ X_1 \end{bmatrix}$$

Aside – Simulation of ARMA Processes: VII

- leads to

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= \phi X_1 + U_2 \\ X_3 &= \phi X_2 + U_3 \\ X_4 &= \phi X_3 + U_4 \\ X_5 &= \phi X_4 + U_5 \\ X_6 &= \phi X_5 + U_6 \end{aligned}$$

- if we were to increase sample size to 1000, would have

$$X_{1000} = \phi X_{999} + U_{1000},$$

which requires a single multiplication

Aside – Simulation of ARMA Processes: VIII

- Q: why not use $\boldsymbol{X}_n = C_n' \boldsymbol{U}_n$ instead, where $C_n = A_n^{-1}$?
  - use IA to get elements of C_n' & variances of innovations
  - simulate innovations $\boldsymbol{U}_n$ (again, easy to do!)
  - no unraveling needed
- consider AR(1) example again:

$$\begin{bmatrix} X_6 \\ X_5 \\ X_4 \\ X_3 \\ X_2 \\ X_1 \end{bmatrix} = \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 \\ 0 & 1 & \phi & \phi^2 & \phi^3 & \phi^4 \\ 0 & 0 & 1 & \phi & \phi^2 & \phi^3 \\ 0 & 0 & 0 & 1 & \phi & \phi^2 \\ 0 & 0 & 0 & 0 & 1 & \phi \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_6 \\ U_5 \\ U_4 \\ U_3 \\ U_2 \\ U_1 \end{bmatrix}$$

Aside – Simulation of ARMA Processes: IX

- leads to

$$
\begin{aligned}
X_1 &= U_1 \\
X_2 &= U_2 + \phi U_1 \\
X_3 &= U_3 + \phi U_2 + \phi^2 U_1 \\
X_4 &= U_4 + \phi U_3 + \phi^2 U_2 + \phi^3 U_1 \\
X_5 &= U_5 + \phi U_4 + \phi^2 U_3 + \phi^3 U_2 + \phi^4 U_1 \\
X_6 &= U_6 + \phi U_5 + \phi^2 U_4 + \phi^3 U_3 + \phi^4 U_2 + \phi^5 U_1
\end{aligned}
$$

- if we were to increase sample size to 1000, would have

$$X_{1000} = U_{1000} + \phi U_{999} + \cdots + \phi^{998} U_2 + \phi^{999} U_1,$$

which requires 999 multiplications as opposed to 1 using L–D approach – hence IA approach is computationally unattractive

Multi-Step-Ahead Prediction: I

- reconsider one-step-ahead predictor $\widehat{X}_{n+1}$ of X_{n+1} given X_n, $X_{n-1}, \ldots, X_1$
- in preparation for considering multi-step-ahead prediction, will now denote $\widehat{X}_{n+1}$ by $\widehat{X}_{n+1|n}$
- $\widehat{X}_{n+1|n}$ can be written as either a linear combination of previous time series values or previous innovations:

$$\widehat{X}_{n+1|n} = \sum_{j=1}^{n} \phi_{n,j} X_{n-j+1} \text{ or } \widehat{X}_{n+1|n} = \sum_{j=1}^{n} \theta_{n,j} U_{n-j+1}$$

- for a given $h \geq 2$, want to formulate best linear predictor $\widehat{X}_{n+h|n}$ of X_{n+h} given $X_n, X_{n-1}, \ldots, X_1$

Multi-Step-Ahead Prediction: II

- first approach: replacing n in

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

with $n + h - 1$ gives

$$\widehat{X}_{n+h|n+h-1} = \sum_{j=1}^{n+h-1} \phi_{n+h-1,j} X_{n+h-j}$$

- above involves unobserved $X_{n+h-1}, \ldots, X_{n+1}$, but replacing these with $\widehat{X}_{n+h-1|n}, \ldots, \widehat{X}_{n+1|n}$ gives desired predictor:

$$\widehat{X}_{n+h|n} = \sum_{j=1}^{h-1} \phi_{n+h-1,j} \widehat{X}_{n+h-j|n} + \sum_{j=h}^{n+h-1} \phi_{n+h-1,j} X_{n+h-j}$$

Multi-Step-Ahead Prediction: III

- leads to recursive scheme for computing $\widehat{X}_{n+h|n}$ starting with one-step-ahead predictor $\widehat{X}_{n+1|n}$ (we know how to get this!)
- two-step-ahead predictor: replace X_{n+1} in

$$\widehat{X}_{n+2|n+1} = \sum_{j=1}^{n+1} \phi_{n+1,j} X_{n+2-j}$$

with $\widehat{X}_{n+1|n}$ to get

$$\widehat{X}_{n+2|n} = \phi_{n+1,1} \widehat{X}_{n+1|n} + \sum_{j=2}^{n+1} \phi_{n+1,j} X_{n+2-j}$$

Multi-Step-Ahead Prediction: IV

- three-step-ahead predictor: replace X_{n+2} & X_{n+1} in

$$\widehat{X}_{n+3|n+2} = \sum_{j=1}^{n+2} \phi_{n+2,j} X_{n+3-j}$$

with $\widehat{X}_{n+2|n}$ & $\widehat{X}_{n+1|n}$ to get

$$\widehat{X}_{n+3|n} = \phi_{n+2,1}\widehat{X}_{n+2|n} + \phi_{n+2,2}\widehat{X}_{n+1|n} + \sum_{j=3}^{n+2} \phi_{n+2,j} X_{n+3-j}$$

- yadda, yadda, yadda, coming eventually to the desired

$$\widehat{X}_{n+h|n} = \sum_{j=1}^{h-1} \phi_{n+h-1,j}\widehat{X}_{n+h-j|n} + \sum_{j=h}^{n+h-1} \phi_{n+h-1,j} X_{n+h-j}$$

Multi-Step-Ahead Prediction: V

- since $\widehat{X}_{n+1|n}, \ldots, \widehat{X}_{n+h-1|n}$ are all linear combinations of X_n, $\ldots, X_1$, it follows that $\widehat{X}_{n+h|n}$ is also such:

$$\begin{aligned}\widehat{X}_{n+h|n} &= \sum_{j=1}^{h-1} \phi_{n+h-1,j}\widehat{X}_{n+h-j|n} + \sum_{j=h}^{n+h-1} \phi_{n+h-1,j}X_{n+h-j} \\ &\stackrel{\text{def}}{=} \sum_{j=1}^{n} a_j X_{n-j+1}\end{aligned}$$

- can show that $\boldsymbol{a}_n = [a_1, \ldots, a_n]'$ so defined is a solution to

$$\Gamma_n \boldsymbol{a}_n = \boldsymbol{\gamma}_n(h),$$

where $n \times n$ matrix Γ_n has (i,j)th entry of $\gamma(i-j)$, while

$$\boldsymbol{\gamma}_n(h) = [\gamma(h), \ldots, \gamma(h+n-1)]'$$

Multi-Step-Ahead Prediction: VI

- second approach: replacing n in

$$\widehat{X}_{n+1|n} = \sum_{j=1}^{n} \theta_{n,j} U_{n-j+1}$$

with $n + h - 1$ gives

$$\widehat{X}_{n+h|n+h-1} = \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} U_{n+h-j}$$

- above involves unobserved $U_{n+h-1}, \ldots, U_{n+1}$, but replacing these with their expected values (zero!) gives desired predictor:

$$\widehat{X}_{n+h|n} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} U_{n+h-j} = \sum_{j=1}^{n} \theta_{n+h-1,n+h-j} U_j$$

Multi-Step-Ahead Prediction: VII

- MSE of h-step-ahead forecast is

$$E\{(X_{n+h} - \widehat{X}_{n+h|n})^2\} = E\{X_{n+h}^2\} - 2E\{X_{n+h}\widehat{X}_{n+h|n}\} + E\{\widehat{X}_{n+h|n}^2\}$$
$$= \gamma(0) - E\{\widehat{X}_{n+h|n}^2\} = \gamma(0) - \operatorname{var}\{\widehat{X}_{n+h|n}\}$$

since $E\{X_{n+h}^2\} = \gamma(0)$ and $E\{X_{n+h}\widehat{X}_{n+h|n}\} = E\{\widehat{X}_{n+h|n}^2\}$ (exercise!)

- since $\operatorname{var}\{U_j\} = v_{j-1}$ and U_j's are uncorrelated,

$$\operatorname{var}\{\widehat{X}_{n+h|n}\} = \operatorname{var}\left\{\sum_{j=1}^{n} \theta_{n+h-1,n+h-j} U_j\right\} = \sum_{j=1}^{n} \theta_{n+h-1,n+h-j}^2 v_{j-1}$$

- MSE is thus given by

$$E\{(X_{n+h} - \widehat{X}_{n+h|n})^2\} = \gamma(0) - \sum_{j=1}^{n} \theta_{n+h-1,n+h-j}^2 v_{j-1} \stackrel{\text{def}}{=} \sigma_n^2(h)$$

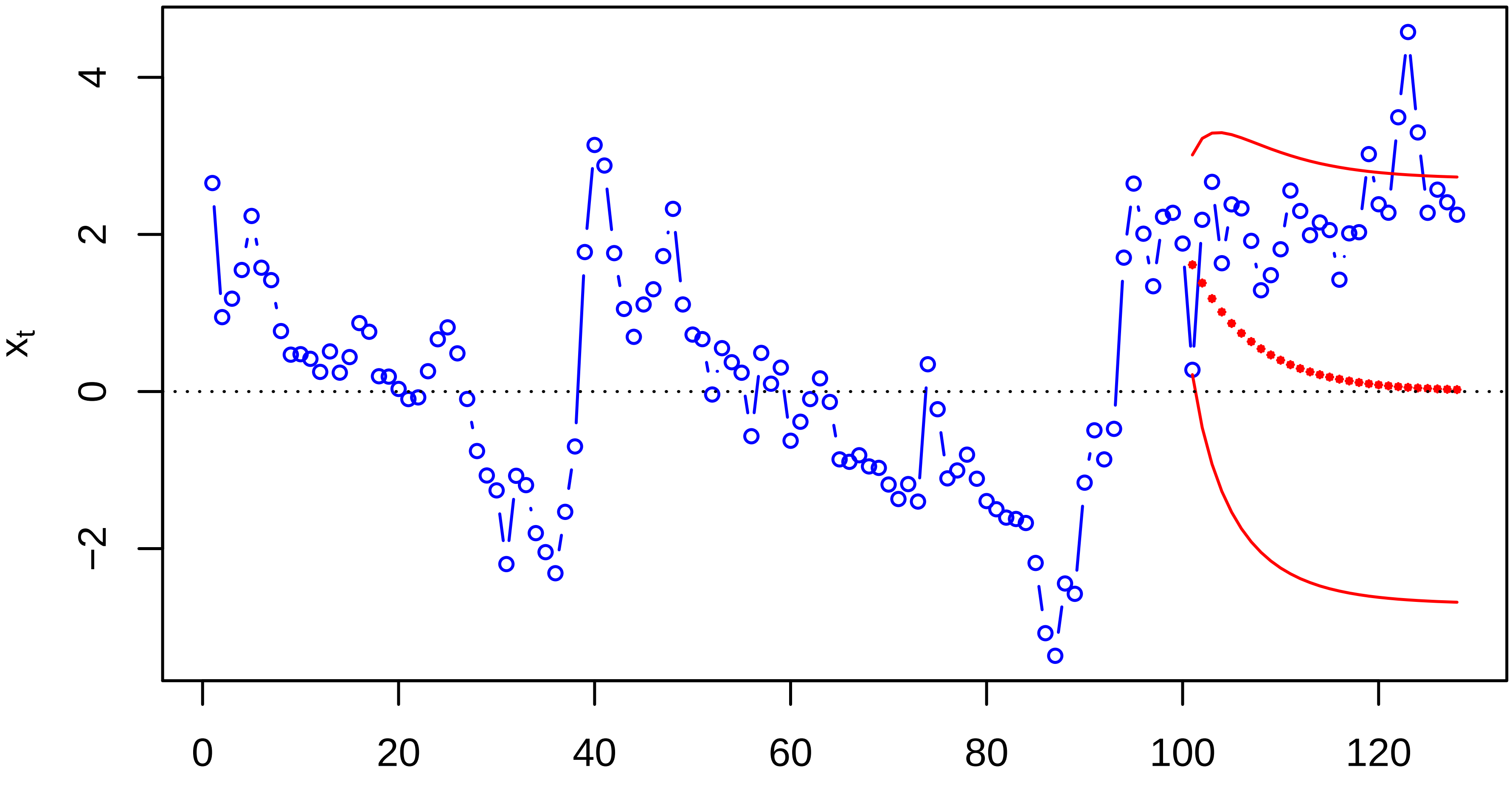
Multi-Step-Ahead Prediction: VIII

- under a Gaussian assumption, can use above to form 95% *prediction bounds* for unknown X_{n+h}:

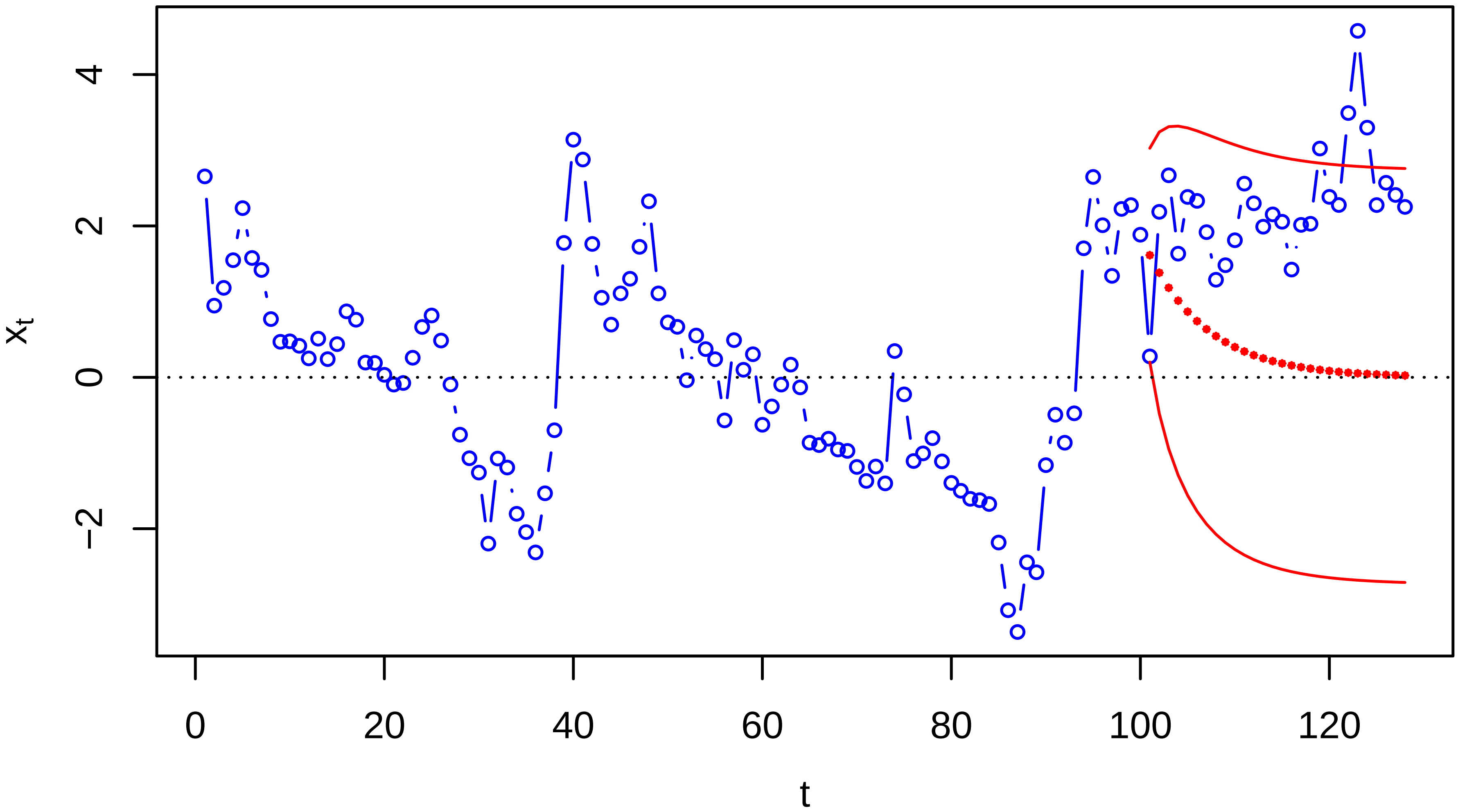
$$\widehat{X}_{n+h|n} \pm 1.96\sigma_n(h)$$

- as example, consider 1st part of wind speed series $x'_1, \ldots, x'_{100}$
- after centering x'_t by subtracting off its sample mean $\bar{x}'$, we model $x_t = x'_t - \bar{x}'$ as an AR(1) process $X_t = \phi X_{t-1} + Z_t$ with ϕ estimated by $\hat{\phi} = \hat{\rho}(1) \doteq 0.856$ (cf. overhead X–16)
- based on $x_1, \ldots, x_{100}$, forecast last 28 values $x'_{101} - \bar{x}'$, ..., $x'_{128} - \bar{x}'$ of time series and see how well we do
- following overheads show results from
  - homegrown `R` code based on theory presented above
  - built-in `R` functions `ar` and `predict`

Multi-Step-Ahead Prediction of Wind Speed



Multi-Step-Ahead Prediction of Wind Speed using R



Predictions Based on Infinite Past: I

- rather than using $X_n, \ldots, X_1$ to predict X_{n+h}, suppose we use, for some $m \geq 0$,
$$X_n, \ldots, X_1, X_0, X_{-1}, \ldots, X_{-m}$$
and form best linear predictor to be denoted by $\widehat{X}_{n+h|n,m}$
- by letting $m \to \infty$ and assuming limit exists (in MS sense), can write
$$\widehat{X}_{n+h|n,\infty} = \sum_{j=1}^{\infty} \alpha_j X_{n-j+1},$$
where α_j's are set by a version of the orthogonality principle:
$$\operatorname{cov}\left\{X_{n+h} - \sum_{j=1}^{\infty} \alpha_j X_{n-j+1}, X_{n-i}\right\} = 0, \quad i = 0, 1, \ldots$$

BD: 65–66, SS: 107

Predictions Based on Infinite Past: II

- refer to $\widehat{X}_{n+h|n,\infty}$ as *predictor of* X_{n+h} *based on infinite past* $X_n, X_{n-1}, \ldots$
- associated prediction error $X_{n+h} - \widehat{X}_{n+h|n,\infty}$ has MSE

$$E\{(X_{n+h} - \widehat{X}_{n+h|n,\infty})^2\} = \text{var}\,\{X_{n+h} - \widehat{X}_{n+h|n,\infty}\},$$

which can be compared to

$$\text{var}\,\{X_{n+h} - \widehat{X}_{n+h|n}\}$$

to see how much can be gained from having lots more data (recall that $\widehat{X}_{n+h|n}$ is based on just $X_n, X_{n-1}, \ldots, X_1$)

Predictions Based on Infinite Past: III

- applying representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ at } t = n+h \text{ yields } X_{n+h} = \sum_{j=0}^{\infty} \psi_j Z_{n+h-j}$$

- consider Z_t's that make up X_{n+h} but *not* X_n; i.e., Z_{n+h}, $Z_{n+h-1}, \ldots, Z_{n+1}$
- replacing these h RVs by their expected values (zero) suggests

$$\widehat{X}_{n+h|n,\infty} = \sum_{j=h}^{\infty} \psi_j Z_{n+h-j}$$

- prediction error is thus

$$X_{n+h} - \widehat{X}_{n+h|n,\infty} = \sum_{j=0}^{\infty} \psi_j Z_{n+h-j} - \sum_{j=h}^{\infty} \psi_j Z_{n+h-j} = \sum_{j=0}^{h-1} \psi_j Z_{n+h-j}$$

Predictions Based on Infinite Past: IV

- note that, in keeping with version of the orthogonality principle appropriate here, for $i = 0, 1, \ldots,$

$$\operatorname{cov}\{X_{n+h} - \widehat{X}_{n+h|n,\infty}, X_{n-i}\} = \operatorname{cov}\left\{\sum_{j=0}^{h-1}\psi_j Z_{n+h-j}, \sum_{j=0}^{\infty}\psi_j Z_{n-i-j}\right\}$$
$$= 0$$

because the first and second sums on the right-hand side do not have any RVs in common: first sum involves Z_{n+1}, Z_{n+2}, $\ldots$, Z_{n+h}, while the second involves a subset of Z_n, Z_{n-1}, $\ldots$

Predictions Based on Infinite Past: V

- since $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, variance of

$$X_{n+h} - \widehat{X}_{n+h|n,\infty} = \sum_{j=0}^{h-1} \psi_j Z_{n+h-j}$$

i.e., MSE of $\widehat{X}_{n+h|n,\infty}$, is given by

$$\mathrm{var}\,\{X_{n+h} - \widehat{X}_{n+h|n,\infty}\} = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2$$

- in particular, for $h = 1$, MSE is

$$\mathrm{var}\,\{X_{n+1} - \widehat{X}_{n+1|n,\infty}\} = \sigma^2 \text{ rather than } v_n = \mathrm{var}\,\{X_{n+1} - \widehat{X}_{n+1|n}\}$$

- exercise: compare variances of $\widehat{X}_{n+1|n,\infty}$ and $\widehat{X}_{n+1}$ for specific MA(1) and AR(1) processes and selected values for n

References

- H. A. David (1985), 'Bias of S^2 Under Dependence,' *The American Statistician*, **39**, p. 201
- J. Durbin (1960), 'The Fitting of Time Series Models,' *Revue de l'Institut International de Statistique/Review of the International Statistical Institute*, **28**, pp. 233–44
- S. M. Kay (1981), 'Efficient Generation of Colored Noise,' *Proceedings of the IEEE*, **69**, pp. 480–1
- N. Levinson (1947), 'The Wiener RMS Error Criterion in Filter Design and Prediction,' *Journal of Mathematical Physics*, **25**, pp. 261–78
- A. Papoulis (1985), 'Levinson's Algorithm, Wold's Decomposition, and Spectral Estimation,' *SIAM Review*, **27**, pp. 405–41