Durbin-Levinson

$$\phi_0 = 0, \qquad \phi_{00} = 0;$$

$$\phi_1 = \phi_{11}, \qquad \phi_{11} = \frac{\gamma(1)}{\gamma(0)};$$

$$\phi_n = \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}, \quad \phi_{nn} = \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.$$

$$\phi_n = (\phi_{n1}, \ldots, \phi_{nn})' \qquad \tilde{\phi}_n = (\phi_{nn}, \ldots, \phi_{n1})',$$

$$\gamma_n = (\gamma(1), \ldots, \gamma(n))' \qquad \tilde{\gamma}_n = (\gamma(n), \ldots, \gamma(1))'.$$

Durbin-Levinson: Example

$$\phi_0 = 0, \qquad \phi_{00} = 0;$$

$$\phi_1 = \phi_{11}, \qquad \phi_{11} = \frac{\gamma(1)}{\gamma(0)};$$

$$\phi_n = \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}, \quad \phi_{nn} = \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.$$

This algorithm computes $\phi_1, \phi_2, \phi_3, \ldots$, where

$$X_2^1 = X_1\phi_1, \quad X_3^2 = (X_2, X_1)\phi_2, \quad X_4^3 = (X_3, X_2, X_1)\phi_3, \ldots$$

Durbin-Levinson: Example

$$\phi_1 = \phi_{11}, \qquad \phi_{11} = \frac{\gamma(1)}{\gamma(0)};$$

$$\phi_n = \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}, \quad \phi_{nn} = \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.$$

$$\phi_1 = \gamma(1)/\gamma(0),$$

$$\phi_2 = \begin{pmatrix} \phi_1 - \phi_{22}\phi_{11} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \frac{\gamma(1)}{\gamma(0)}\left(1 - \frac{\gamma(2)-\gamma(1)}{\gamma(0)-\gamma(1)}\right) \\ \frac{\gamma(2)-\gamma(1)}{\gamma(0)-\gamma(1)} \end{pmatrix}, \text{ etc.}$$

The innovations representation

Instead of writing the best linear predictor as

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

we can write

$$X_{n+1}^n = \theta_{n1} \underbrace{\left(X_n - X_n^{n-1}\right)}_{\text{innovation}} + \theta_{n2} \left(X_{n-1} - X_{n-1}^{n-2}\right) + \cdots + \theta_{nn} \left(X_1 - X_1^0\right).$$

This is still linear in $X_1, \ldots, X_n$.

The innovations are uncorrelated:
$\mathrm{Cov}(X_j - X_j^{j-1}, X_i - X_i^{i-1}) = 0$ for $i \neq j$.

Comparing representations: $U_n = X_n - X_n^{n-1}$ versus X_n

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

$$\begin{pmatrix} X_1^0 \\ X_2^1 \\ \vdots \\ X_n^{n-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & & 0 \\ \vdots & & \ddots & \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \cdots & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

Innovations Algorithm

$$X_1^0 = 0, \qquad X_{n+1}^n = \sum_{i=1}^{n} \theta_{ni} \left(X_{n+1-i} - X_{n+1-i}^{n-i} \right).$$

NB: error in text.

$$\theta_{n,n-i} = \frac{1}{P_{i+1}^i} \left(\gamma(n-i) - \sum_{j=0}^{i-1} \theta_{i,i-j} \theta_{n,n-j} P_{j+1}^j \right).$$

$$P_1^0 = \gamma(0) \qquad P_{n+1}^n = \gamma(0) - \sum_{i=0}^{n-1} \theta_{n,n-i}^2 P_{i+1}^i.$$

Innovations Algorithm: Example

$$\theta_{n,n-i} = \frac{1}{P_{i+1}^{i}} \left(\gamma(n-i) - \sum_{j=0}^{i-1} \theta_{i,i-j} \theta_{n,n-j} P_{j+1}^{j} \right).$$

$$P_1^0 = \gamma(0) \qquad P_{n+1}^{n} = \gamma(0) - \sum_{i=0}^{n-1} \theta_{n,n-i}^2 P_{i+1}^{i}.$$

$$\theta_{1,1} = \gamma(1)/P_1^0, \qquad P_2^1 = \gamma(0) - \theta_{1,1}^2 P_1^0$$

$$\theta_{2,2} = \gamma(2)/P_1^0, \quad \theta_{2,1} = \left(\gamma(1) - \theta_{1,1}\theta_{2,2}P_1^0\right)/P_2^1,$$

$$P_3^2 = \gamma(0) - \left(\theta_{2,2}^2 P_1^0 + \theta_{2,1}^2 P_2^1\right)$$

$$\theta_{3,3}, \quad \theta_{3,2}, \quad \theta_{3,1}, \quad P_4^3, \ldots$$

Predicting h steps ahead using innovations

The innovations representation for the one-step-ahead forecast is

$$P(X_{n+1}|X_1,\ldots,X_n) = \sum_{i=1}^{n} \theta_{ni} \left(X_{n+1-i} - X_{n+1-i}^{n-i}\right),$$

What is the innovations representation for $P(X_{n+h}|X_1,\ldots,X_n)$?

Fact: If $h \geq 1$ and $1 \leq i \leq n$, we have
$\mathrm{Cov}(X_{n+h} - P(X_{n+h}|X_1,\ldots,X_{n+h-1}), X_i) = 0.$

Thus, $P(X_{n+h} - P(X_{n+h}|X_1,\ldots,X_{n+h-1})|X_1,\ldots,X_n) = 0.$
That is, the best prediction of X_{n+h} is the
best prediction of the one-step-ahead forecast of X_{n+h}.

Predicting h steps ahead using innovations

$$P(X_{n+h}|X_1, \ldots, X_n)$$

$$= P\left(P(X_{n+h}|X_1, \ldots, X_{n+h-1})|X_1, \ldots, X_n\right)$$

$$= P\left(\sum_{i=1}^{n+h-1} \theta_{n+h-1,i}\left(X_{n+h-i} - X_{n+h-i}^{n+h-i+1}\right)|X_1, \ldots, X_n\right)$$

$$= \sum_{i=1}^{n+h-1} \theta_{n+h-1,i} P\left(\left(X_{n+h-i} - X_{n+h-i}^{n+h-i+1}\right)|X_1, \ldots, X_n\right)$$

$$= \sum_{i=h}^{n+h-1} \theta_{n+h-1,i} P\left(\left(X_{n+h-i} - X_{n+h-i}^{n+h-i+1}\right)|X_1, \ldots, X_n\right)$$

$$= \sum_{i=h}^{n+h-1} \theta_{n+h-1,i}\left(X_{n+h-i} - X_{n+h-i}^{n+h-i+1}\right)$$

Predicting h steps ahead using innovations

$$P(X_{n+1}|X_1,\ldots,X_n) = \sum_{i=1}^{n} \theta_{ni} \left(X_{n+1-i} - X_{n+1-i}^{n-i}\right)$$

$$P(X_{n+h}|X_1,\ldots,X_n) = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left(X_{n+h-j} - X_{n+h-j}^{n+h-j+1}\right)$$

$$= \sum_{i=1}^{n} \theta_{n+h-1,h-1+i} \left(X_{n+1-i} - X_{n+1-i}^{n-i}\right)$$

$(j = i + h - 1)$

Mean squared error of h-step-ahead forecasts

From orthogonality of the predictors and the error,
$\mathrm{E}\left((X_{n+h} - P(X_{n+h}|X_1,\ldots,X_n))\, P(X_{n+h}|X_1,\ldots,X_n)\right) = 0.$

That is, $\mathrm{E}\left(X_{n+h}P(X_{n+h}|X_1,\ldots,X_n)\right) = \mathrm{E}\left(P(X_{n+h}|X_1,\ldots,X_n)^2\right)$.

Hence, we can express the mean squared error as

$$\begin{aligned} P_{n+h}^n &= \mathrm{E}\left(X_{n+h} - P(X_{n+h}|X_1,\ldots,X_n)\right)^2 \\ &= \gamma(0) + \mathrm{E}\left(P(X_{n+h}|X_1,\ldots,X_n)\right)^2 \\ &\qquad - 2\mathrm{E}\left(X_{n+h}P(X_{n+h}|X_1,\ldots,X_n)\right) \\ &= \gamma(0) - \mathrm{E}\left(P(X_{n+h}|X_1,\ldots,X_n)\right)^2. \end{aligned}$$

Mean squared error of h-step-ahead forecasts

But the innovations are uncorrelated, so

$$\begin{aligned}
P_{n+h}^{n} &= \gamma(0) - \mathrm{E}\left(P(X_{n+h}|X_1,\ldots,X_n)\right)^2 \\
&= \gamma(0) - \mathrm{E}\left(\sum_{j=h}^{n+h-1} \theta_{n+h-1,j}\left(X_{n+h-j} - X_{n+h-j}^{n+h-j-1}\right)\right)^2 \\
&= \gamma(0) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2\, \mathrm{E}\left(X_{n+h-j} - X_{n+h-j}^{n+h-j-1}\right)^2 \\
&= \gamma(0) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2\, P_{n+h-j}^{n+h-j-1}.
\end{aligned}$$

Example: Innovations algorithm for forecasting an MA(1)

Suppose that we have an MA(1) process $\{X_t\}$ satisfying

$$X_t = W_t + \theta_1 W_{t-1}.$$

Given $X_1, X_2, \ldots, X_n$, we wish to compute the best linear forecast of X_{n+1}, using the innovations representation,

$$X_1^0 = 0, \qquad X_{n+1}^n = \sum_{i=1}^{n} \theta_{ni} \left(X_{n+1-i} - X_{n+1-i}^{n-i} \right).$$

Example: Innovations algorithm for forecasting an MA(1)

An aside: The linear predictions are in the form

$$X_{n+1}^n = \sum_{i=1}^{n} \theta_{ni} Z_{n+1-i}$$

for uncorrelated, zero mean random variables Z_i. In particular,

$$X_{n+1} = Z_{n+1} + \sum_{i=1}^{n} \theta_{ni} Z_{n+1-i},$$

where $Z_{n+1} = X_{n+1} - X_{n+1}^n$ (and all the Z_i are uncorrelated).
This is suggestive of an MA representation. Why isn't it an MA?

Example: Innovations algorithm for forecasting an MA(1)

$$\theta_{n,n-i} = \frac{1}{P_{i+1}^{i}} \left(\gamma(n-i) - \sum_{j=0}^{i-1} \theta_{i,i-j} \theta_{n,n-j} P_{j+1}^{j} \right).$$

$$P_1^0 = \gamma(0) \qquad P_{n+1}^{n} = \gamma(0) - \sum_{i=0}^{n-1} \theta_{n,n-i}^2 P_{i+1}^{i}.$$

The algorithm computes $P_1^0 = \gamma(0), \theta_{1,1}$ (in terms of $\gamma(1)$); $P_2^1, \theta_{2,2}$ (in terms of $\gamma(2)$), $\theta_{2,1}$; $P_3^2, \theta_{3,3}$ (in terms of $\gamma(3)$), etc.

Example: Innovations algorithm for forecasting an MA(1)

$$\theta_{n,n-i} = \frac{1}{P_{i+1}^{i}} \left(\gamma(n-i) - \sum_{j=0}^{i-1} \theta_{i,i-j}\theta_{n,n-j}P_{j+1}^{j} \right).$$

For an MA(1), $\gamma(0) = \sigma^2(1+\theta_1^2), \gamma(1) = \theta_1\sigma^2$.

Thus: $\theta_{1,1} = \gamma(1)/P_1^0$;

$\theta_{2,2} = 0, \theta_{2,1} = \gamma(1)/P_2^1$;

$\theta_{3,3} = \theta_{3,2} = 0; \theta_{3,1} = \gamma(1)/P_3^2$, etc.

Because $\gamma(n-i) \neq 0$ only for $i = n-1$, only $\theta_{n,1} \neq 0$.

Example: Innovations algorithm for forecasting an MA(1)

For the MA(1) process $\{X_t\}$ satisfying

$$X_t = W_t + \theta_1 W_{t-1},$$

the innovations representation of the best linear forecast is

$$X_1^0 = 0, \qquad X_{n+1}^n = \theta_{n1}\left(X_n - X_n^{n-1}\right).$$

More generally, for an MA(q) process, we have $\theta_{ni} = 0$ for $i > q$.

Example: Innovations algorithm for forecasting an MA(1)

For the MA(1) process $\{X_t\}$,

$$X_1^0 = 0, \qquad X_{n+1}^n = \theta_{n1}\left(X_n - X_n^{n-1}\right).$$

This is consistent with the observation that

$$X_{n+1} = Z_{n+1} + \sum_{i=1}^{n} \theta_{ni} Z_{n+1-i},$$

where the uncorrelated Z_i are defined by $Z_t = X_t - X_t^{t-1}$ for $t = 1, \ldots, n+1$.

Indeed, as n increases, $P_{n+1}^n \to \mathrm{Var}(W_t)$ (recall the recursion for P_{n+1}^n), and $\theta_{n1} = \gamma(1)/P_n^{n-1} \to \theta_1$.

Recall: Forecasting an AR(p)

For the AR(p) process $\{X_t\}$ satisfying

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + W_t,$$

$$X_1^0 = 0, \qquad X_{n+1}^n = \sum_{i=1}^{p} \phi_i X_{n+1-i}$$

for $n \geq p$. Then

$$X_{n+1} = \sum_{i=1}^{p} \phi_i X_{n+1-i} + Z_{n+1},$$

where $Z_{n+1} = X_{n+1} - X_{n+1}^n$.

The Durbin-Levinson algorithm is convenient for AR(p) processes.

The innovations algorithm is convenient for MA(q) processes.

Linear prediction based on the infinite past

So far, we have considered linear predictors based on n observed values of the time series:

$$X_{n+m}^n = P(X_{n+m}|X_n, X_{n-1}, \ldots, X_1).$$

What if we have access to *all* previous values, $X_n, X_{n-1}, X_{n-2}, \ldots$?

Write

$$\begin{aligned}\tilde{X}_{n+m} &= P(X_{n+m}|X_n, X_{n-1}, \ldots) \\ &= \sum_{i=1}^{\infty} \alpha_i X_{n+1-i}.\end{aligned}$$

Linear prediction based on the infinite past

$$\tilde{X}_{n+m} = P(X_{n+m}|X_n, X_{n-1}, \ldots) = \sum_{i=1}^{\infty} \alpha_i X_{n+1-i}.$$

The orthogonality property of the optimal linear predictor implies

$$\mathrm{E}\left[(\tilde{X}_{n+m} - X_{n+m})X_{n+1-i}\right] = 0, \quad i = 1, 2, \ldots$$

Thus, if $\{X_t\}$ is a zero-mean stationary time series, we have

$$\sum_{j=1}^{\infty} \alpha_j \gamma(i-j) = \gamma(m-1+i), \quad i = 1, 2, \ldots$$

Linear prediction based on the infinite past

If $\{X_t\}$ is a causal, invertible, *linear* process, we can write

$$X_{n+m} = \sum_{j=1}^{\infty} \psi_j W_{n+m-j} + W_{n+m}, \quad W_{n+m} = \sum_{j=1}^{\infty} \pi_j X_{n+m-j} + X_{n+m}.$$

In this case,

$$\begin{aligned}
\tilde{X}_{n+m} &= P(X_{n+m}|X_n, X_{n-1}, \ldots) \\
&= P(W_{n+m}|X_n, \ldots) - \sum_{j=1}^{\infty} \pi_j P(X_{n+m-j}|X_n, \ldots) \\
&= -\sum_{j=1}^{m-1} \pi_j P(X_{n+m-j}|X_n, \ldots) - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}.
\end{aligned}$$

Linear prediction based on the infinite past

$$\tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j P(X_{n+m-j} | X_n, \ldots) - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}.$$

That is, $\tilde{X}_{n+1} = -\sum_{j=1}^{\infty} \pi_j X_{n+1-j}$,

$$\tilde{X}_{n+2} = -\pi_1 \tilde{X}_{n+1} - \sum_{j=2}^{\infty} \pi_j X_{n+2-j},$$

$$\tilde{X}_{n+3} = -\pi_1 \tilde{X}_{n+2} - \pi_2 \tilde{X}_{n+1} - \sum_{j=3}^{\infty} \pi_j X_{n+3-j}.$$

The invertible (AR(∞)) representation gives the forecasts $\tilde{X}_{n+m}^{n}$.

Linear prediction based on the infinite past

To compute the mean squared error, we notice that

$$\begin{aligned}\tilde{X}_{n+m} = P(X_{n+m}|X_n, X_{n-1}, \ldots) &= \sum_{j=1}^{\infty} \psi_j P(W_{n+m-j}|X_n, X_{n-1}, \ldots) \\ &\quad + P(W_{n+m}|X_n, X_{n-1}, \ldots) \\ &= \sum_{j=m}^{\infty} \psi_j W_{n+m-j}.\end{aligned}$$

$$\begin{aligned}\mathrm{E}\left(X_{n+m} - P(X_{n+m}|X_n, X_{n-1}, \ldots)\right)^2 &= \mathrm{E}\left(\sum_{j=0}^{m-1} \psi_j W_{n+m-j}\right)^2 \\ &= \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2.\end{aligned}$$

Linear prediction based on the infinite past

That is, the mean squared error of the forecast based on the infinite history is given by the initial terms of the causal (MA(∞)) representation:

$$\mathrm{E}\left(X_{n+m} - \tilde{X}_{n+m}\right)^2 = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2.$$

In particular, for $m = 1$, the mean squared error is σ_w^2.

The truncated forecast

For large n, truncating the infinite-past forecasts gives a good approximation:

$$\tilde{X}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j}$$

$$\tilde{X}^n_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{X}^n_{n+m-j} - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j}.$$

The approximation is exact for AR(p) when $n \geq p$, since $\pi_j = 0$ for $j > p$. In general, it is a good approximation if the π_j converge quickly to 0.

Example: Forecasting an ARMA(p,q) model

Consider an ARMA(p,q) model:

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = W_t + \sum_{i=1}^{q} \theta_i W_{t-i}.$$

Suppose we have $X_1, X_2, \ldots, X_n$, and we wish to forecast X_{n+m}.

We could use the best linear prediction, X_{n+m}^n.

For an AR(p) model (that is, $q = 0$), we can write down the coefficients ϕ_n.

Otherwise, we must solve a linear system of size n.

If n is large, the truncated forecasts $\tilde{X}_{n+m}^n$ give a good approximation. To compute them, we could compute π_i and truncate.

There is also a recursive method, which takes time $O((n+m)(p+q))$...

Recursive truncated forecasts for an ARMA(p,q) model

$$\tilde{W}_t^n = 0 \quad \text{for } t \le 0. \qquad \tilde{X}_t^n = \begin{cases} 0 & \text{for } t \le 0, \\ X_t & \text{for } 1 \le t \le n. \end{cases}$$

$$\tilde{W}_t^n = \tilde{X}_t^n - \phi_1 \tilde{X}_{t-1}^n - \cdots - \phi_p \tilde{X}_{t-p}^n$$

$$- \theta_1 \tilde{W}_{t-1}^n - \cdots - \theta_q \tilde{W}_{t-q}^n \qquad \text{for } t = 1, \ldots, n.$$

$$\tilde{W}_t^n = 0 \qquad \text{for } t > n.$$

$$\tilde{X}_t^n = \phi_1 \tilde{X}_{t-1}^n + \cdots + \phi_p \tilde{X}_{t-p}^n + \theta_1 \tilde{W}_{t-1}^n + \cdots + \theta_q \tilde{W}_{t-q}^n$$

$$\text{for } t = n+1, \ldots, n+m.$$

Example: Forecasting an AR(2) model

Consider the following AR(2) model.

$$X_t + \frac{1}{1.21} X_{t-2} = W_t.$$

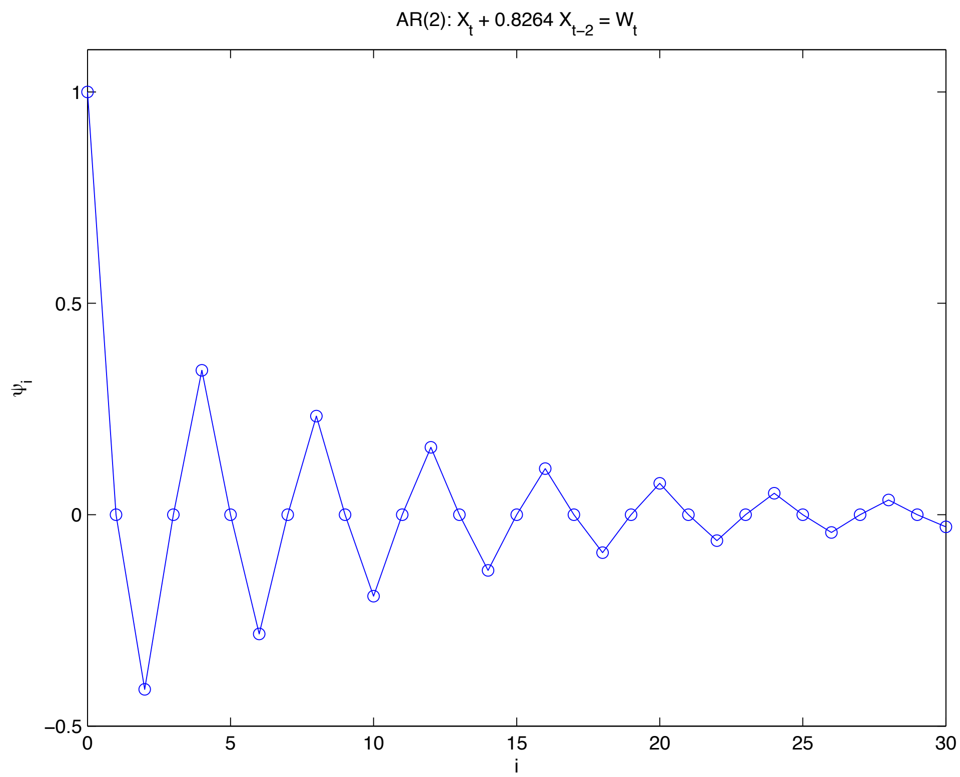
The zeros of the characteristic polynomial $z^2 + 1.21$ are at $\pm 1.1i$. We can solve the linear difference equations $\psi_0 = 1$, $\phi(B)\psi_t = 0$ to compute the MA(∞) representation:

$$\psi_t = \frac{1}{2} 1.1^{-t} \cos(\pi t/2).$$

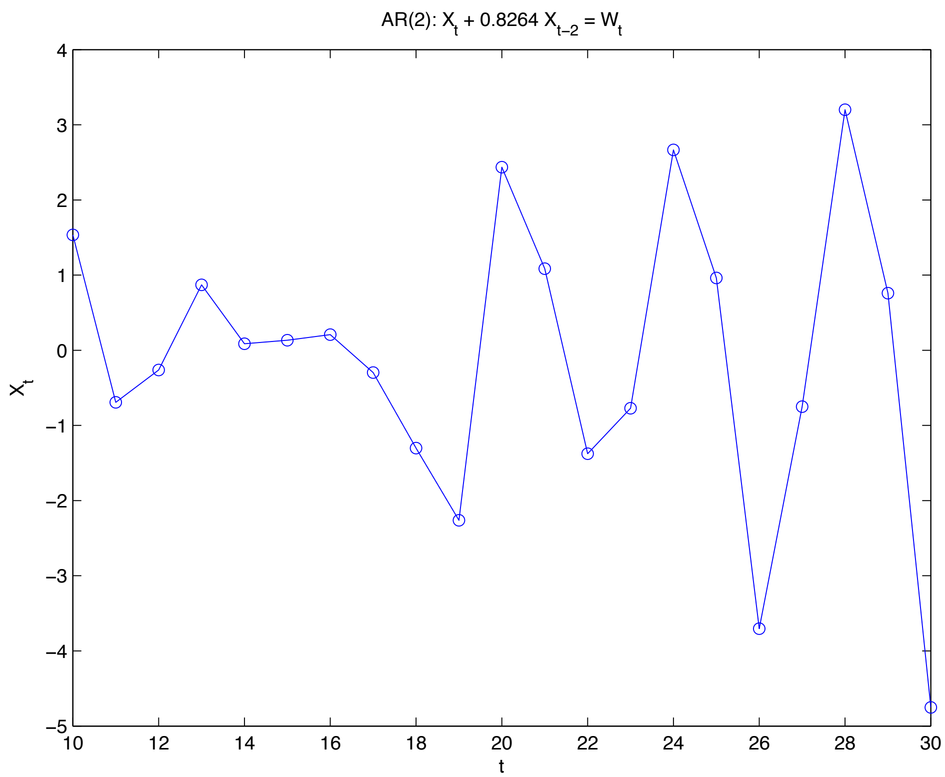
Thus, the m-step-ahead estimates have mean squared error

$$\mathrm{E}(X_{n+m} - \tilde{X}_{n+m})^2 = \sum_{j=0}^{m-1} \psi_j^2.$$

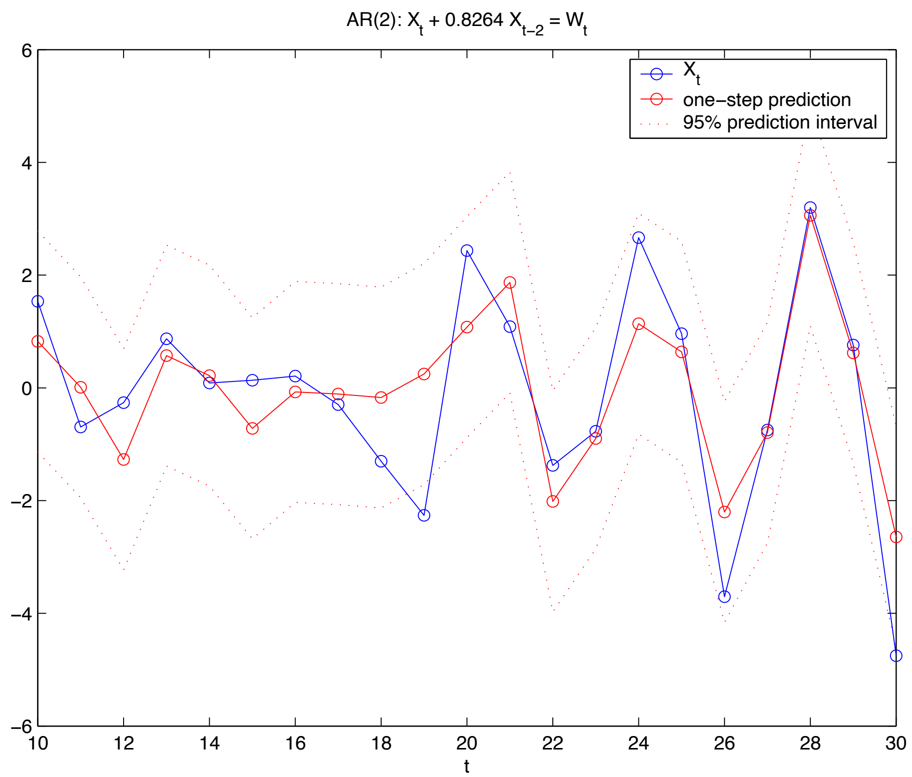
Example: Forecasting an AR(2) model



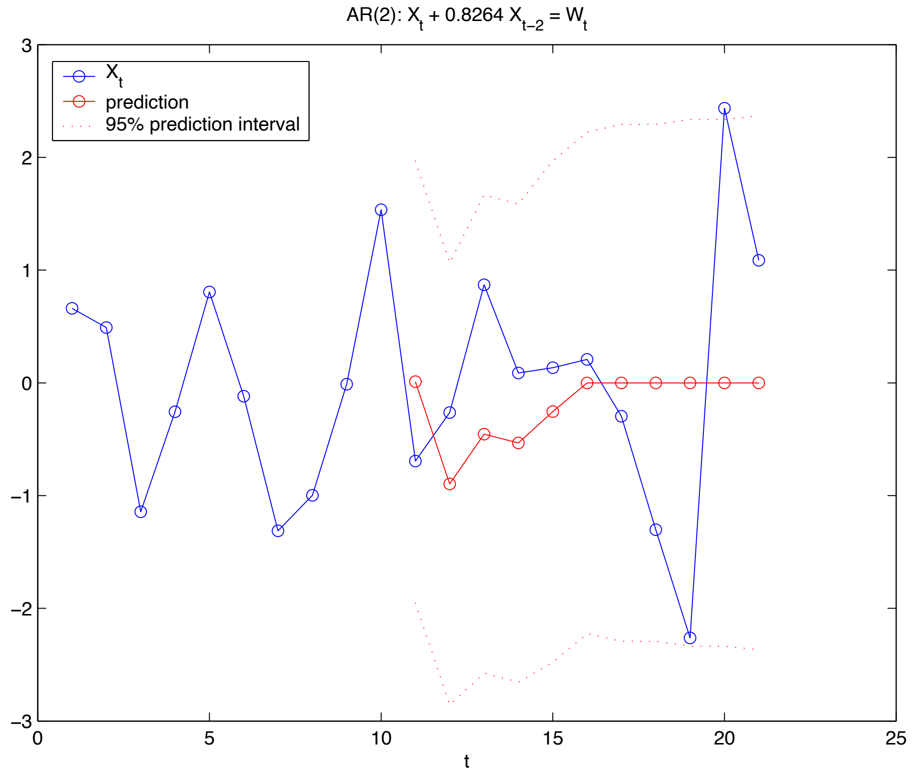
Example: Forecasting an AR(2) model



Example: Forecasting an AR(2) model



Example: Forecasting an AR(2) model



Review (Lecture 1): Time series modelling and forecasting

1. Plot the time series.
   Look for trends, seasonal components, step changes, outliers.
2. Transform data so that residuals are **stationary**.
   (a) Remove trend and seasonal components.
   (b) Differencing.
   (c) Nonlinear transformations (log, $\sqrt{\cdot}$).
3. Fit model to residuals.
4. Forecast time series by forecasting residuals and inverting any transformations.

Review: Time series modelling and forecasting

Stationary time series models: ARMA(p,q).

- $p = 0$: MA(q),
- $q = 0$: AR(p).

We have seen that any causal, invertible linear process has:
an MA(∞) representation (from causality), and
an AR(∞) representation (from invertibility).

Real data cannot be *exactly* modelled using a finite number of parameters.

We choose p, q to give a simple but accurate model.

Review: Time series modelling and forecasting

How do we use data to decide on p, q?

1. Use sample ACF/PACF to make preliminary choices of model order.
2. Estimate parameters for each of these choices.
3. Compare predictive accuracy/complexity of each (using, e.g., AIC).

NB: We need to compute parameter estimates for several different model orders.

Thus, recursive algorithms for parameter estimation are important.

We'll see that some of these are identical to the recursive algorithms for forecasting.

Review: Time series modelling and forecasting

Model:	**ACF:**	**PACF:**
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA(p,q)	decays	decays

Parameter estimation

We want to estimate the parameters of an ARMA(p,q) model.
We will assume (for now) that:
1. The model order (p and q) is known, and
2. The data has zero mean.

If (2) is not a reasonable assumption, we can subtract the sample mean $\bar{y}$, fit a zero-mean ARMA model,

$$\phi(B)X_t = \theta(B)W_t,$$

to the mean-corrected time series $X_t = Y_t - \bar{y}$,
and then use $X_t + \bar{y}$ as the model for Y_t.

Parameter estimation: Maximum likelihood estimator

One approach:

Assume that $\{X_t\}$ is Gaussian, that is, $\phi(B)X_t = \theta(B)W_t$, where W_t is i.i.d. Gaussian.

Choose ϕ_i, θ_j to maximize the *likelihood*:

$$L(\phi, \theta, \sigma^2) = f(X_1, \ldots, X_n),$$

where f is the joint (Gaussian) density for the given ARMA model.

(c.f. choosing the parameters that maximize the probability of the data.)

Parameter estimation: Maximum likelihood estimator

Advantages of MLE:

Efficient (low variance estimates).

Often the Gaussian assumption is reasonable.

Even if $\{X_t\}$ is not Gaussian, the asymptotic distribution of the estimates $(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2)$ is the same as the Gaussian case.

Disadvantages of MLE:

Difficult optimization problem.

Need to choose a good starting point (often use other estimators for this).

Preliminary parameter estimates

Yule-Walker for AR(p): Regress X_t onto $X_{t-1}, \ldots, X_{t-p}$.

Durbin-Levinson algorithm with γ replaced by $\hat{\gamma}$.

Yule-Walker for ARMA(p,q): Method of moments. Not efficient.

Innovations algorithm for MA(q): with γ replaced by $\hat{\gamma}$.

Hannan-Rissanen algorithm for ARMA(p,q):

1. Estimate high-order AR.
2. Use to estimate (unobserved) noise W_t.
3. Regress X_t onto $X_{t-1}, \ldots, X_{t-p}, \hat{W}_{t-1}, \ldots, \hat{W}_{t-q}$.
4. Regress again with improved estimates of W_t.