Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

Last lecture:

1. Recursive forecasting method: Durbin-Levinson.
2. The innovations representation.
3. Recursive method: Innovations algorithm.

Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

1. Review: Forecasting, the innovations representation.
2. Forecasting h steps ahead.
3. Example: Innovations algorithm for forecasting an MA(1)
4. An aside: Innovations algorithm for forecasting an ARMA(p,q)
5. Linear prediction based on the infinite past
6. The truncated predictor

Review: Forecasting

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

$$\Gamma_n \phi_n = \gamma_n,$$

$$P_{n+1}^n = \mathrm{E}\left(X_{n+1} - X_{n+1}^n\right)^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n,$$

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

$$\phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn})', \quad \gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))'.$$

Review: Partial autocorrelation function

The Partial AutoCorrelation Function (PACF) of a stationary time series $\{X_t\}$ is

$$\phi_{11} = \mathrm{Corr}(X_1, X_0) = \rho(1)$$

$$\phi_{hh} = \mathrm{Corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}) \quad \text{for } h = 2, 3, \ldots$$

This removes the linear effects of $X_1, \ldots, X_{h-1}$:

$$\ldots, X_{-1}, \underline{X_0}, \underbrace{X_1, X_2, \ldots, X_{h-1}}_{\text{partial out}}, \underline{X_h}, X_{h+1}, \ldots$$

Review: Partial autocorrelation function

The PACF ϕ_{hh} is also the last coefficient in the best linear prediction of X_{h+1} given $X_1, \ldots, X_h$:

$$\Gamma_h \phi_h = \gamma_h \qquad X_{h+1}^h = \phi_h' X$$
$$\phi_h = (\phi_{h1}, \phi_{h2}, \ldots, \phi_{hh}).$$

Prediction error variance reduces by a factor $1 - \phi_{nn}^2$:

$$\begin{aligned} P_{n+1}^n &= \gamma(0) - \phi_n' \gamma_n \\ &= P_n^{n-1}\left(1 - \phi_{nn}^2\right). \end{aligned}$$

Review: 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:
$\text{Cov}(X_j - X_j^{j-1}, X_i - X_i^{i-1}) = 0$ for $i \neq j$.

Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

1. Review: Forecasting, the innovations representation.
2. Forecasting h steps ahead.
3. Example: Innovations algorithm for forecasting an MA(1)

Predicting h steps ahead using innovations

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}.

Fact: The best prediction of $X_{n+1} - X_{n+1}^n$ given only $X_1, \ldots, X_n$ is 0.

Similarly for $n + 2, \ldots, n + h - 1$.

Predicting h steps ahead using innovations

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

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

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

Mean squared error of h-step-ahead forecasts

But the innovations are uncorrelated, so

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

Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

1. Review: Forecasting, the innovations representation.
2. Forecasting h steps ahead.
3. Example: Innovations algorithm for forecasting an MA(1)

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.

Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

1. Review: Forecasting, the innovations representation.
2. Forecasting h steps ahead.
3. Example: Innovations algorithm for forecasting an MA(1)
4. An aside: Innovations algorithm for forecasting an ARMA(p,q)
5. Linear prediction based on the infinite past
6. The truncated predictor

An aside: Forecasting an ARMA(p,q)

There is a related representation for an ARMA(p,q) process, based on the innovations algorithm. Suppose that $\{X_t\}$ is an ARMA(p,q) process:

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + W_t + \sum_{j=1}^{q} \theta_j W_{t-j}.$$

Consider the transformed process (C. F. Ansley, Biometrika 66: 59–65, 1979)

$$Z_t = \begin{cases} X_t/\sigma & \text{if } t = 1, \ldots, p, \\ \phi(B) X_t/\sigma & \text{if } t > p. \end{cases}$$

If $p > 0$, this is not stationary. However, there is a more general version of the innovations algorithm, which is applicable to nonstationary processes.

An aside: Forecasting an ARMA(p,q)

Let $\theta_{n,j}$ be the coefficients obtained from the application of the innovations algorithm to this process Z_t. This gives the representation

$$X_{n+1}^n = \begin{cases} \sum_{j=1}^n \theta_{nj} \left(X_{n+1-j} - X_{n+1-j}^{n-j} \right) & n < p, \\ \sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{nj} \left(X_{n+1-j} - X_{n+1-j}^{n-j} \right) & n \geq p \end{cases}$$

For a causal, invertible $\{X_t\}$:
$\mathrm{E}(X_n - X_n^{n-1} - W_n)^2 \to 0$, $\theta_{nj} \to \theta_j$, and $P_n^{n+1} \to \sigma^2$.

Notice that this illustrates one way to simulate an ARMA(p,q) process exactly.

Why?

Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

1. Review: Forecasting, the innovations representation.
2. Forecasting h steps ahead.
3. Example: Innovations algorithm for forecasting an MA(1)
4. An aside: Innovations algorithm for forecasting an ARMA(p,q)
5. Linear prediction based on the infinite past
6. The truncated predictor

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 $\tilde{X}^n_{n+m}$.

Linear prediction based on the infinite past

To compute the mean squared error, we notice that

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

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

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:

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

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

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_{n+m}.

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 $\tilde{X}^n_{n+m}$ 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

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

$$\begin{aligned} \tilde{W}_t^n = \tilde{X}_t^n - \phi_1 \tilde{X}_{t-1}^n - \cdots - \phi_p \tilde{X}_{t-p}^n \\ - \theta_1 \tilde{W}_{t-1}^n - \cdots - \theta_q \tilde{W}_{t-q}^n \qquad \text{for } t = 1, \ldots, n. \end{aligned}$$

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

$$\tilde{X}_t^n = \phi_1 \tilde{X}_{t-1}^n + \cdots + \phi_p \tilde{X}_{t-p}^n + \theta_1 \tilde{W}_{t-1}^n + \cdots + \theta_q \tilde{W}_{t-q}^n$$
$$\text{for } t = n+1, \ldots, n+m.$$

Example: Forecasting an AR(2) model

Consider the following AR(2) model.

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

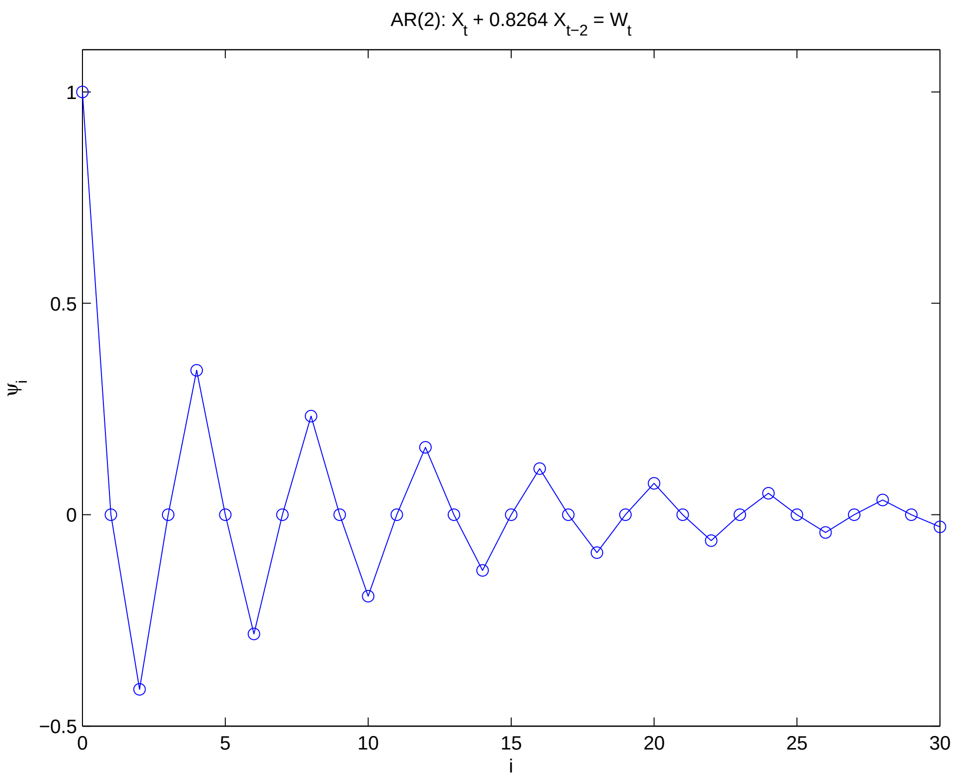
The zeros of the characteristic polynomial $z^2 + 1.21$ are at $\pm 1.1i$. We can solve the linear difference equations $\psi_0 = 1$, $\phi(B)\psi_t = 0$ to compute the MA(∞) representation:

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

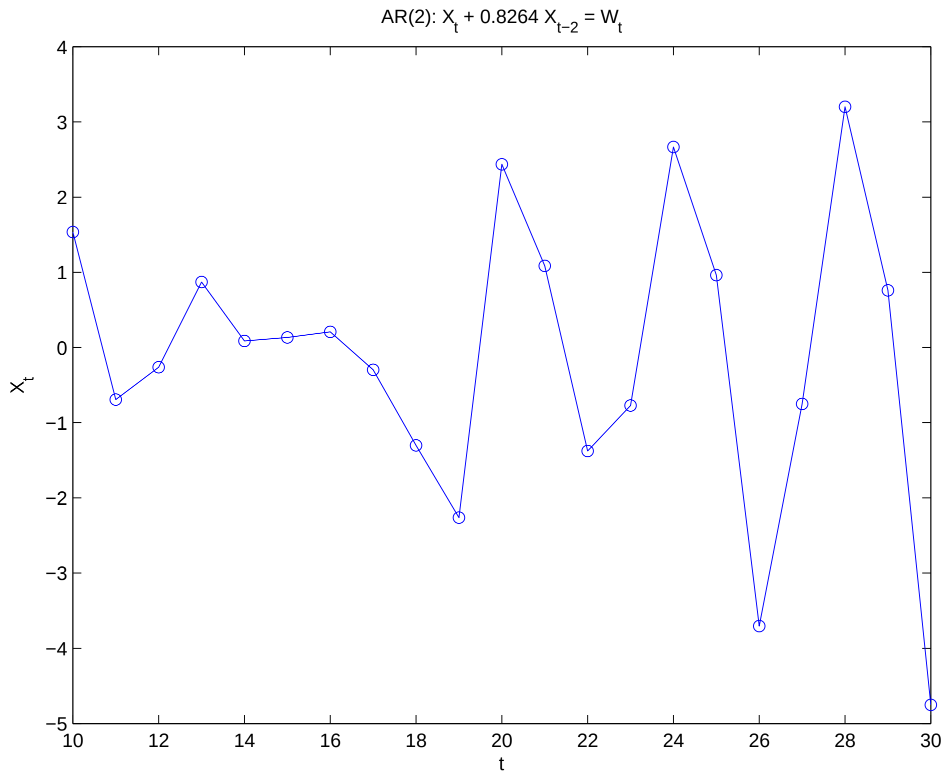
Thus, the m-step-ahead estimates have mean squared error

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

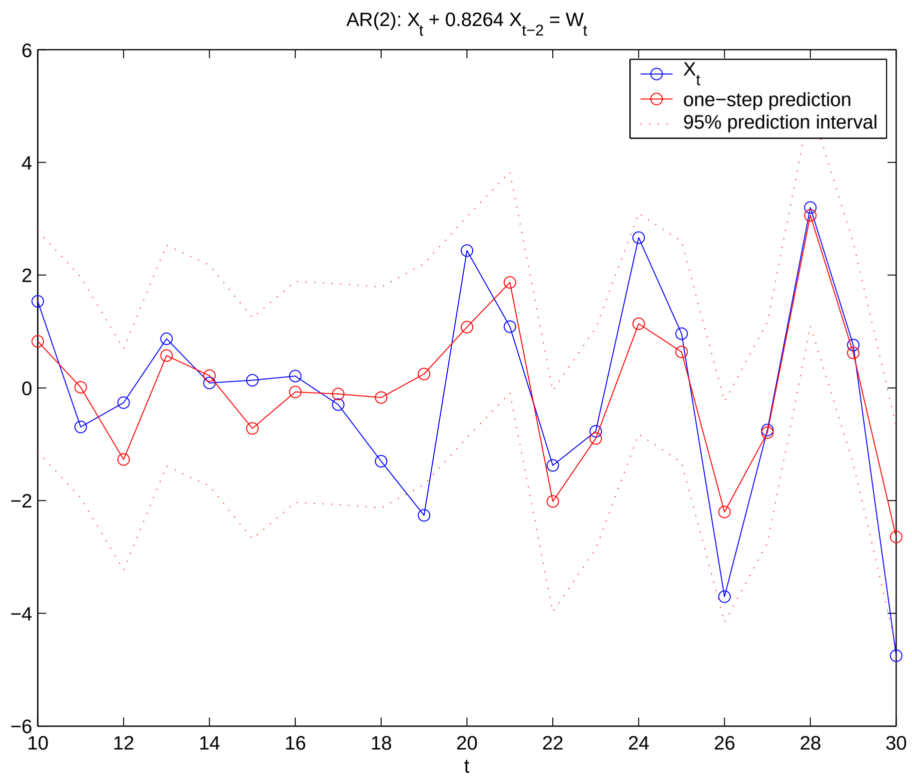
Example: Forecasting an AR(2) model



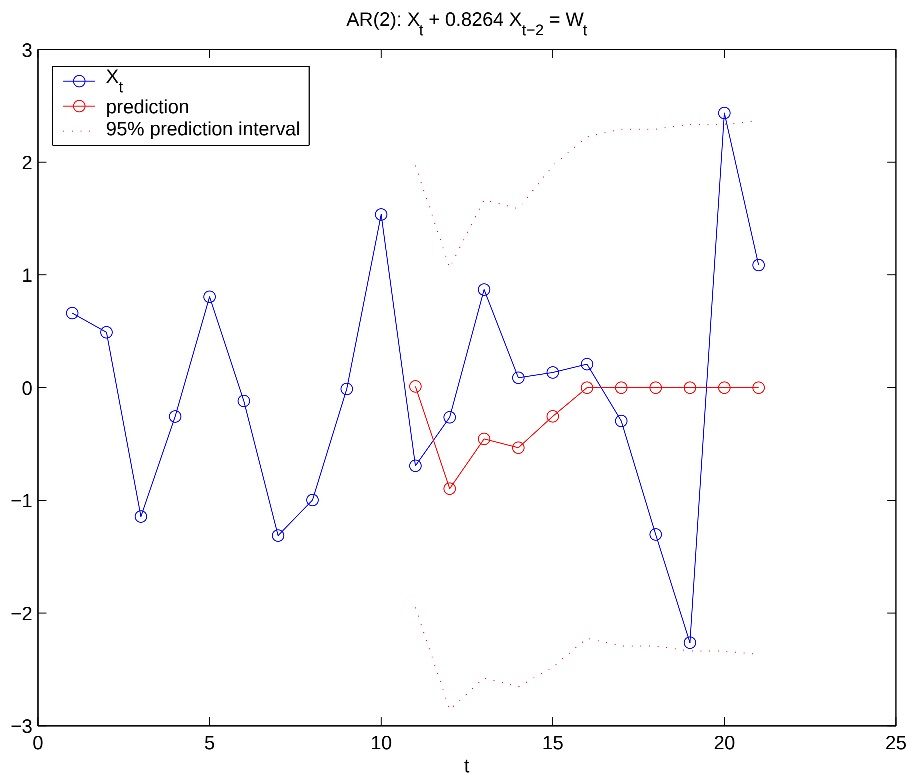
Example: Forecasting an AR(2) model



Example: Forecasting an AR(2) model



Example: Forecasting an AR(2) model



Introduction to Time Series Analysis. Lecture 10.

Peter Bartlett

1. Review: Forecasting, the innovations representation.
2. Forecasting h steps ahead.
3. Example: Innovations algorithm for forecasting an MA(1)
4. An aside: Innovations algorithm for forecasting an ARMA(p,q)
5. Linear prediction based on the infinite past
6. The truncated predictor