Last lecture:

1. Forecasting and backcasting.
2. Prediction operator.
3. Partial autocorrelation function.
Peter Bartlett

1. Review: Forecasting
2. Partial autocorrelation function.
4. The innovations representation.
6. Example: Innovations algorithm for forecasting an MA(1)
Review: One-step-ahead linear prediction

\[ X_{n+1}^n = \phi_{n1}X_n + \phi_{n2}X_{n-1} + \cdots + \phi_{nn}X_1 \]
\[ \Gamma_n \phi_n = \gamma_n, \]
\[ P_{n+1}^n = E \left( X_{n+1} - X_{n+1}^n \right)^2 = \gamma(0) - \gamma' \Gamma_n^{-1} \gamma_n, \]
\[ \Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}, \]
\[ \phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn})', \quad \gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))'. \]
Review: The prediction operator

For random variables $Y, Z_1, \ldots, Z_n$, define the **best linear prediction of $Y$ given $Z = (Z_1, \ldots, Z_n)'** as the operator $P(\cdot | Z)$ applied to $Y$:

$$
P(Y|Z) = \mu_Y + \phi'(Z - \mu_Z)
$$

with

$$
\Gamma \phi = \gamma,
$$

where

$$
\gamma = \text{Cov}(Y, Z)
$$

$$
\Gamma = \text{Cov}(Z, Z).
$$
Review: Properties of the prediction operator

1. $E(Y - P(Y|Z)) = 0$, $E((Y - P(Y|Z))Z) = 0$.
2. $E((Y - P(Y|Z))^2) = \text{Var}(Y) - \phi'\gamma$.
3. $P(\alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_0|Z) = \alpha_0 + \alpha_1 P(Y_1|Z) + \alpha_2 P(Y_2|Z)$.
4. $P(Z_i|Z) = Z_i$.
5. $P(Y|Z) = EY$ if $\gamma = 0$. 
Review: Partial autocorrelation function

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

\[
\phi_{11} = \text{Corr}(X_1, X_0) = \rho(1)
\]

\[
\phi_{hh} = \text{Corr}(X_h - X_{h-1}^h, X_0 - X_{0}^{h-1}) \text{ for } h = 2, 3, \ldots
\]

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

\[
\ldots, X_{-1}, \underbrace{X_0, X_1, X_2, \ldots, X_{h-1}}_{\text{partial out}}, X_h, X_{h+1}, \ldots
\]
The PACF $\phi_{hh}$ is also the last coefficient in the best linear prediction of $X_{h+1}$ given $X_1, \ldots, X_h$:

\[
\Gamma_h \phi_h = \gamma_h \quad X_{h+1}^h = \phi'_h X \\
\phi_h = (\phi_{h1}, \phi_{h2}, \ldots, \phi_{hh}).
\]
Example: PACF of an AR(p)

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

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

Thus, $\phi_{hh} = \begin{cases} \phi_h & \text{if } 1 \leq h \leq p \\ 0 & \text{otherwise.} \end{cases}$
Example: PACF of an invertible MA(q)

For \( X_t = \sum_{i=1}^{q} \theta_i W_{t-i} + W_t, \quad X_t = -\sum_{i=1}^{\infty} \pi_i X_{t-i} + W_t, \)

\[
X_{n+1}^{n} = P(X_{n+1}|X_1, \ldots, X_n) = P \left( \sum_{i=1}^{\infty} \pi_i X_{n+1-i} + W_{n+1}|X_1, \ldots, X_n \right)
= -\sum_{i=1}^{\infty} \pi_i P(X_{n+1-i}|X_1, \ldots, X_n)
= -\sum_{i=1}^{n} \pi_i X_{n+1-i} - \sum_{i=n+1}^{\infty} \pi_i P(X_{n+1-i}|X_1, \ldots, X_n).
\]

In general, \( \phi_{hh} \neq 0. \)
ACF of the MA(1) process

MA(1): \( X_t = Z_t + \theta Z_{t-1} \)

\[ \frac{\theta}{1 + \theta^2} \]
ACF of the AR(1) process

AR(1): $X_t = \phi X_{t-1} + Z_t$

The diagram shows the autocorrelation function (ACF) of the AR(1) process. The x-axis represents lags $\phi^{|h|}$, and the y-axis shows the autocorrelation values. The graph indicates that the autocorrelation decays exponentially with increasing lag.
PACF of the MA(1) process

MA(1): $X_t = Z_t + \theta Z_{t-1}$
PACF of the AR(1) process

AR(1): $X_t = \phi X_{t-1} + Z_t$
<table>
<thead>
<tr>
<th>Model</th>
<th>ACF:</th>
<th>PACF:</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(p)</td>
<td>decays</td>
<td>zero for $h &gt; p$</td>
</tr>
<tr>
<td>MA(q)</td>
<td>zero for $h &gt; q$</td>
<td>decays</td>
</tr>
<tr>
<td>ARMA(p,q)</td>
<td>decays</td>
<td>decays</td>
</tr>
</tbody>
</table>
For a realization $x_1, \ldots, x_n$ of a time series, the **sample PACF** is defined by

$$\hat{\phi}_{00} = 1$$

$$\hat{\phi}_{hh} = \text{last component of } \hat{\phi}_h,$$

where $\hat{\phi}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h$. 

**Sample PACF**

15

1. Review: Forecasting
2. Partial autocorrelation function.
4. The innovations representation.
6. Example: Innovations algorithm for forecasting an MA(1)
The importance of $P_{n+1}^n$: Prediction intervals

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

\[ \Gamma_n \phi_n = \gamma_n, \quad P_{n+1}^n = E \left( X_{n+1} - X_{n+1}^n \right)^2 = \gamma(0) - \gamma' \Gamma_n^{-1} \gamma_n. \]

After seeing $X_1, \ldots, X_n$, we forecast $X_{n+1}^n$. The expected squared error of our forecast is $P_{n+1}^n$. We can construct a prediction interval:

\[ X_{n+1}^n \pm c_{\alpha/2} \sqrt{P_{n+1}^n}. \]

For a Gaussian process, the prediction error has distribution $\mathcal{N}(0, P_{n+1}^n)$, so $c_{0.05/2} = 1.96$ gives a 95% prediction interval.
Computing linear prediction coefficients

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

\[ \Gamma_n \phi_n = \gamma_n, \]

\[ P_{n+1}^n = \mathbb{E} (X_{n+1}^n - X_{n+1}^n)^2 = \gamma(0) - \gamma' \Gamma_n^{-1} \gamma_n. \]

How can we compute these quantities recursively? i.e., given the coefficients \( \phi_{n-1} \) of \( X_{n-1}^n \), how can we compute the coefficients \( \phi_n \) of \( X_{n+1}^n \), without solving another linear system \( \Gamma_n \phi_n = \gamma_n \)?
\[ \begin{align*}
\phi_0 &= 0, & \phi_{00} &= 0; \\
\phi_1 &= \phi_{11}, & \phi_{11} &= \frac{\gamma(1)}{\gamma(0)}; \\
\phi_n &= \begin{pmatrix}
\phi_{n-1} - \phi_{nn} \tilde{\phi}_{n-1} \\
\phi_{nn}
\end{pmatrix}, & \phi_{nn} &= \frac{\gamma(n) - \phi'_{n-1} \tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1} \gamma_{n-1}}. \\
\phi_n &= (\phi_{n1}, \ldots, \phi_{nn})', & \tilde{\phi}_n &= (\phi_{nn}, \ldots, \phi_{11})', \\
\gamma_n &= (\gamma(1), \ldots, \gamma(n))', & \tilde{\gamma}_n &= (\gamma(n), \ldots, \gamma(1))'.
\end{align*} \]
Durbin-Levinson: Example

\[ \begin{align*}
\phi_0 &= 0, & \phi_{00} &= 0; \\
\phi_1 &= \phi_{11}, & \phi_{11} &= \frac{\gamma(1)}{\gamma(0)}; \\
\phi_n &= \begin{pmatrix}
\phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\
\phi_{nn}
\end{pmatrix}, & \phi_{nn} &= \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.
\end{align*} \]

This algorithm computes \( \phi_1, \phi_2, \phi_3, \ldots \), where

\[ \begin{align*}
X_2^1 &= X_1\phi_1, & X_3^2 &= (X_2, X_1)\phi_2, & X_4^3 &= (X_3, X_2, X_1)\phi_3, \ldots
\end{align*} \]
Durbin-Levinson: Example

\[ \phi_1 = \phi_{11}, \quad \phi_{11} = \frac{\gamma(1)}{\gamma(0)}; \]

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

\[ \phi_1 = \gamma(1)/\gamma(0), \]

\[ \phi_2 = \begin{pmatrix} \phi_1 - \phi_{22} \phi_{11} \\ \phi_{22} \end{pmatrix} = \left( \begin{array}{c} \frac{\gamma(1)}{\gamma(0)} \left( 1 - \frac{\gamma(2) - \gamma(1)}{\gamma(0) - \gamma(1)} \right) \\ \frac{\gamma(2) - \gamma(1)}{\gamma(0) - \gamma(1)} \end{array} \right), \text{ etc.} \]
Clearly, $\Gamma_1 \phi_1 = \gamma_1$.

Suppose $\Gamma_{n-1} \phi_{n-1} = \gamma_{n-1}$. Then $\Gamma_{n-1} \tilde{\phi}_{n-1} = \tilde{\gamma}_{n-1}$, and so

$$
\Gamma_n \phi_n = \begin{pmatrix} \Gamma_{n-1} & \tilde{\gamma}_{n-1} \\ \tilde{\gamma}_{n-1}' & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n-1} - \phi_{nn} \tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}
= \begin{pmatrix} \gamma_{n-1} \\ \tilde{\gamma}_{n-1}' \phi_{n-1} + \phi_{nn} (\gamma(0) - \gamma'_{n-1} \phi_{n-1}) \end{pmatrix}
= \gamma_n.
$$
\[ P_{n+1}^n = \gamma(0) - \phi'_n \gamma_n \]
\[ = \gamma(0) - \begin{pmatrix} \phi_{n-1} - \phi_{nn} \tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}' \begin{pmatrix} \gamma_{n-1} \\ \gamma(n) \end{pmatrix} \]
\[ = P_{n}^{n-1} - \phi_{nn} \left( \gamma(n) - \tilde{\phi}'_{n-1} \gamma_{n-1} \right) \]
\[ = P_{n}^{n-1} - \phi_{nn}^2 \left( \gamma(0) - \phi'_{n-1} \gamma_{n-1} \right) \quad \text{(From expression for } \phi_{nn} \text{)} \]
\[ = P_{n}^{n-1} \left( 1 - \phi_{nn}^2 \right) . \]

i.e., variance reduces by a factor \( 1 - \phi_{nn}^2 \).

1. Review: Forecasting
2. Partial autocorrelation function.
4. The innovations representation.
6. Example: Innovations algorithm for forecasting an MA(1)
The innovations representation

Instead of writing the best linear predictor as

\[ X_{n+1}^n = \phi_1 X_n + \phi_2 X_{n-1} + \cdots + \phi_n X_1, \]

we can write

\[ X_{n+1}^n = \theta_1 \left( X_n - X_{n-1}^{n-1} \right) + \theta_2 \left( X_{n-1} - X_{n-2}^{n-2} \right) + \cdots + \theta_n \left( X_1 - X_0^0 \right). \]

This is still linear in \( X_1, \ldots, X_n. \)

The innovations are uncorrelated:
\[ \text{Cov}(X_j - X_j^{j-1}, X_i - X_i^{i-1}) = 0 \text{ for } i \neq j. \]
Comparing representations: \( U_n = X_n - X_{n-1} \) versus \( X_n \)

\( \{U_n\} \) form a *decorrelated* representation for the \( \{X_n\} \):

\[
\begin{pmatrix}
U_1 \\
U_2 \\
\vdots \\
U_n
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
-\phi_{11} & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-\phi_{n-1,n-1} & -\phi_{n-1,n-2} & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{pmatrix}
\]
Comparing representations: $U_n = X_n - X_{n-1}^{n-1}$ versus $X_n$

\[
\begin{pmatrix}
X_0 \\
X_1 \\
X_2 \\
\vdots \\
X_{n-1}
\end{pmatrix}
= 
\begin{pmatrix}
0 & 0 & \cdots & 0 \\
\theta_{11} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{n-1,n-1} & \theta_{n-1,n-2} & \cdots & 0
\end{pmatrix}
\begin{pmatrix}
U_1 \\
U_2 \\
\vdots \\
U_n
\end{pmatrix}
\]

1. Review: Forecasting
2. Partial autocorrelation function.
4. The innovations representation.
6. Example: Innovations algorithm for forecasting an MA(1)
Innovations Algorithm

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

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

\[ P_1^0 = \gamma(0) \quad P_{n+1}^n = \gamma(0) - \sum_{i=0}^{n-1} \theta_{n,n-i}^2 P_{i+1}^i. \]
Innovations Algorithm: Example

\[
\theta_{n,n-i} = \frac{1}{P^i_{i+1}} \left( \gamma(n - i) - \sum_{j=0}^{i-1} \theta_{i-i-j} \theta_{n,n-j} P^j_{j+1} \right).
\]

\[
P^0_1 = \gamma(0) \quad P^n_{n+1} = \gamma(0) - \sum_{i=0}^{n-1} \theta^2_{n,n-i} P^i_{i+1}.
\]

\[
\theta_{1,1} = \gamma(1)/P^0_1, \quad P^1_2 = \gamma(0) - \theta^2_{1,1} P^0_1
\]
\[
\theta_{2,2} = \gamma(2)/P^0_1, \quad \theta_{2,1} = (\gamma(1) - \theta_{1,1} \theta_{2,2} P^0_1) / P^1_2,
\]
\[
P^2_3 = \gamma(0) - (\theta^2_{2,2} P^0_1 + \theta^2_{2,1} P^1_2)
\]
\[
\theta_{3,3}, \quad \theta_{3,2}, \quad \theta_{3,1}, \quad P^3_4, \ldots
\]
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 \( P(X_{n+h}|X_1, \ldots, X_n) \)?

It is \( P(X_{n+h}|X_1, \ldots, X_{n+h-1}) \), but with the unobserved innovations (from \( n + 1 \) to \( n + h - 1 \)) set to zero.
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

\[
Cov(X_{n+h} - P(X_{n+h}|X_1, \ldots, X_{n+h-1}), X_i) = 0.
\]

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 $X_{n+1} - X^n_{n+1}$ given only $X_1, \ldots, X_n$ is 0. Similarly for $n + 2, \ldots, n + h - 1$. 
Predicting $h$ steps ahead using innovations

Innovations representation:

$$P(X_{n+h}|X_1, \ldots, X_n) = \sum_{i=1}^{n} \theta_{n+h-1,h-1+i} (X_{n+1-i} - X_{n+1-i}^{n-i})$$
Predicting $h$ steps ahead using innovations (Details)

\[
P(X_{n+h}|X_1, \ldots, X_n)
= P \left( \sum_{i=1}^{n+h-1} \theta_{n+h-1,i} (X_{n+h-i} - X_{n+h-i-1}^{n+h-i}) \mid X_1, \ldots, X_n \right)
= \sum_{i=h}^{n+h-1} \theta_{n+h-1,i} P \left( (X_{n+h-i} - X_{n+h-i}^{n+h-i}) \mid X_1, \ldots, X_n \right)
= \sum_{i=h}^{n+h-1} \theta_{n+h-1,i} (X_{n+h-i} - X_{n+h-i}^{n+h-i-1})
\]
Predicting $h$ steps ahead using innovations (Details)

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

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

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

\[(j = i + h - 1)\]
Mean squared error of $h$-step-ahead forecasts

From orthogonality of the predictors and the error,
\[ E ((X_{n+h} - P(X_{n+h}|X_1, \ldots, X_n)) P(X_{n+h}|X_1, \ldots, X_n)) = 0. \]
That is, \[ E (X_{n+h} P(X_{n+h}|X_1, \ldots, X_n)) = E (P(X_{n+h}|X_1, \ldots, X_n)^2). \]
Hence, we can express the mean squared error as
\[
P_{n+h}^n = E (X_{n+h} - P(X_{n+h}|X_1, \ldots, X_n))^2 \\
= \gamma(0) + E (P(X_{n+h}|X_1, \ldots, X_n))^2 \\
- 2E (X_{n+h} P(X_{n+h}|X_1, \ldots, X_n)) \\
= \gamma(0) - E (P(X_{n+h}|X_1, \ldots, X_n))^2.
\]
But the innovations are uncorrelated, so

\[ P_{n+h}^n = \gamma(0) - E \left( P(X_{n+h}|X_1, \ldots, X_n) \right)^2 \]

\[ = \gamma(0) - \left( \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} \left( X_{n+h-j} - X_{n+h-j-1}^{n+h-j-1} \right) \right)^2 \]

\[ = \gamma(0) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 E \left( X_{n+h-j} - X_{n+h-j}^{n+h-j-1} \right)^2 \]

\[ = \gamma(0) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 \left( X_{n+h-j} - X_{n+h-j}^{n+h-j-1} \right)^2 \]

\[ = \gamma(0) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 P_{n+h-j-1}^{n+h-j-1} \].

1. Review: Forecasting
2. Partial autocorrelation function.
4. The innovations representation.
6. 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,

\[
X_1^0 = 0, \quad X_{n+1}^n = \sum_{i=1}^{n} \theta_{ni} (X_{n+1-i} - X_{n+1-i}^{n-i}).
\]
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 \( Z_{n+1} = X_{n+1} - X_n \) (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)

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

\[
P_1^0 = \gamma(0) \quad P_n^{n+1} = \gamma(0) - \sum_{i=0}^{n-1} \theta_{n,n-i}^2 P_{i+1}^i.
\]

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)

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

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

\[ X_1^0 = 0, \quad X_{n+1}^n = \theta_{n1} (X_n - X_{n-1}^{n-1}). \]

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 \( \{X_t\} \),

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

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 \( Z_t = X_t - X_{t-1}^t \) for \( t = 1, \ldots, n + 1 \).

Indeed, as \( n \) increases, \( P_{n+1}^n \rightarrow \text{Var}(W_t) \) (recall the recursion for \( P_{n+1}^n \)), and \( \theta_{n1} = \gamma(1)/P_n^{n-1} \rightarrow \theta_1 \).
Recall: Forecasting an AR(p)

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

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

\[
X_1^0 = 0, \quad X_{n+1}^n = \sum_{i=1}^{p} \phi_i X_{n+1-i}
\]

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.

1. Review: Forecasting
2. Partial autocorrelation function.
4. The innovations representation.
6. Example: Innovations algorithm for forecasting an MA(1)