Introduction to Time Series Analysis. Lecture 14.

Last lecture: Maximum likelihood estimation

1. Review: Maximum likelihood estimation
2. Model selection
3. Integrated ARMA models
4. Seasonal ARMA
5. Seasonal ARIMA models
Recall: Maximum likelihood estimation

The MLE \((\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2)\) satisfies

\[
\hat{\sigma}_w^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n},
\]

and \(\hat{\phi}, \hat{\theta}\) minimize

\[
\log \left( \frac{S(\hat{\phi}, \hat{\theta})}{n} \right) + \frac{1}{n} \sum_{i=1}^{n} \log r_{i}^{i-1},
\]

where \(r_{i}^{i-1} = P_{i}^{i-1} / \sigma_w^2\) and

\[
S(\phi, \theta) = \sum_{i=1}^{n} \frac{(X_i - X_{i-1})^2}{r_{i}^{i-1}}.
\]
Recall: Maximum likelihood estimation

We can express the likelihood in terms of the innovations. Since the innovations are linear in previous and current values, we can write

\[
\begin{pmatrix}
X_1 \\
\vdots \\
X_n
\end{pmatrix} =
\begin{pmatrix}
C
\end{pmatrix}
\begin{pmatrix}
X_1 - X_1^0 \\
\vdots \\
X_n - X_n^{n-1}
\end{pmatrix}
\]

where \( C \) is a lower triangular matrix with ones on the diagonal. Take the variance/covariance of both sides to see that

\[
\Gamma_n = CDC'^t \quad \text{where} \quad D = \text{diag}(P_1^0, \ldots, P_n^{n-1}).
\]
Recall: Maximum likelihood estimation

\[ |\Gamma_n| = |C|^{2} P_1^0 \cdots P_{n-1}^n = P_1^0 \cdots P_{n-1}^n \] and

\[ X'\Gamma_n^{-1} X = U'C'\Gamma_n^{-1}CU = U'C'C^{-T}D^{-1}C^{-1}CU = U'D^{-1}U. \]

We rewrite the likelihood as

\[
L(\phi, \theta, \sigma_w^2) = \frac{1}{((2\pi)^n P_1^0 \cdots P_{n-1}^n)^{1/2}} \exp \left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(X_i - X_{i-1}^i)^2}{P_{i-1}^i}\right)
\]

\[
= \frac{1}{((2\pi\sigma_w^2)^n r_1^0 \cdots r_{n-1}^n)^{1/2}} \exp \left(-\frac{S(\phi, \theta)}{2\sigma_w^2}\right),
\]

where \(r_{i-1}^i = P_{i-1}^i / \sigma_w^2\) and

\[
S(\phi, \theta) = \sum_{i=1}^{n} \frac{(X_i - X_{i-1}^i)^2}{r_{i}^{i-1}}.
\]
Recall: Maximum likelihood estimation

The log likelihood of $\phi, \theta, \sigma^2_w$ is

$$l(\phi, \theta, \sigma^2_w) = \log(L(\phi, \theta, \sigma^2_w))$$

$$= -\frac{n}{2} \log(2\pi \sigma^2_w) - \frac{1}{2} \sum_{i=1}^{n} \log r_i^{i-1} - \frac{S(\phi, \theta)}{2\sigma^2_w}.$$ 

Differentiating with respect to $\sigma^2_w$ shows that the MLE $(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2_w)$ satisfies

$$\frac{n}{2\hat{\sigma}^2_w} = \frac{S(\hat{\phi}, \hat{\theta})}{2\hat{\sigma}^4_w} \quad \text{⇔} \quad \hat{\sigma}^2_w = \frac{S(\hat{\phi}, \hat{\theta})}{n},$$

and $\hat{\phi}, \hat{\theta}$ minimize

$$\log \left( \frac{S(\hat{\phi}, \hat{\theta})}{n} \right) + \frac{1}{n} \sum_{i=1}^{n} \log r_i^{i-1}.$$
The MLE ($\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2$) satisfies

$$\hat{\sigma}_w^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n},$$

and $\hat{\phi}, \hat{\theta}$ minimize

$$\log \left( \frac{S(\hat{\phi}, \hat{\theta})}{n} \right) + \frac{1}{n} \sum_{i=1}^{n} \log r_i^{i-1},$$

where $r_i^{i-1} = P_i^{i-1}/\sigma_w^2$ and

$$S(\phi, \theta) = \sum_{i=1}^{n} \frac{(X_i - X_i^{i-1})^2}{r_i^{i-1}}.$$
Introduction to Time Series Analysis. Lecture 14.

1. Review: Maximum likelihood estimation
2. Model selection
3. Integrated ARMA models
4. Seasonal ARMA
5. Seasonal ARIMA models
Building ARMA models

1. Plot the time series.
   Look for trends, seasonal components, step changes, outliers.
2. Nonlinearly transform data, if necessary
3. Identify preliminary values of $p$, and $q$.
4. Estimate parameters.
5. Use diagnostics to confirm residuals are white/iid/normal.
6. **Model selection**: Choose $p$ and $q$. 
We have used the data $x$ to estimate parameters of several models. They all fit well (the innovations are white). We need to choose a single model to retain for forecasting. How do we do it?

If we had access to independent data $y$ from the same process, we could compare the likelihood on the new data, $L_y(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2)$.

We could obtain $y$ by leaving out some of the data from our model-building, and reserving it for model selection. This is called *cross-validation*. It suffers from the drawback that we are not using all of the data for parameter estimation.
Model Selection: AIC

We can approximate the likelihood defined using independent data: asymptotically

\[- \ln L_y(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2) \approx - \ln L_x(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2) + \frac{(p + q + 1)n}{n - p - q - 2}.\]

AIC\(_c\): corrected Akaike information criterion.

Notice that:

- More parameters incur a bigger penalty.
- Minimizing the criterion over all values of \(p, q, \hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2\) corresponds to choosing the optimal \(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2\) for each \(p, q\), and then comparing the penalized likelihoods.

There are also other criteria: BIC.
Introduction to Time Series Analysis. Lecture 14.

1. Review: Maximum likelihood estimation
2. Computational simplifications: un/conditional least squares
3. Diagnostics
4. Model selection
5. Integrated ARMA models
6. Seasonal ARMA
7. Seasonal ARIMA models
Integrated ARMA Models: ARIMA(p,d,q)

For \( p, d, q \geq 0 \), we say that a time series \( \{X_t\} \) is an **ARIMA (p,d,q) process** if \( Y_t = \nabla^d X_t = (1 - B)^d X_t \) is ARMA(p,q). We can write

\[
\phi(B)(1 - B)^d X_t = \theta(B) W_t.
\]

Recall the random walk: \( X_t = X_{t-1} + W_t \).

\( X_t \) is not stationary, but \( Y_t = (1 - B)X_t = W_t \) is a stationary process. In this case, it is white, so \( \{X_t\} \) is an ARIMA(0,1,0).

Also, if \( X_t \) contains a trend component plus a stationary process, its first difference is stationary.
Suppose \( \{ X_t \} \) is an ARIMA(0,1,1): \( X_t = X_{t-1} + W_t - \theta_1 W_{t-1} \).

If \( |\theta_1| < 1 \), we can show

\[
X_t = \sum_{j=1}^{\infty} (1 - \theta_1) \theta_1^{j-1} X_{t-j} + W_t,
\]

and so

\[
\tilde{X}_{n+1} = \sum_{j=1}^{\infty} (1 - \theta_1) \theta_1^{j-1} X_{n+1-j}
\]

\[
= (1 - \theta_1) X_n + \sum_{j=2}^{\infty} (1 - \theta_1) \theta_1^{j-1} X_{n+1-j}
\]

\[
= (1 - \theta_1) X_n + \theta_1 \tilde{X}_n.
\]

Exponentially weighted moving average.
Introduction to Time Series Analysis. Lecture 14.

1. Review: Maximum likelihood estimation
2. Computational simplifications: un/conditional least squares
3. Diagnostics
4. Model selection
5. Integrated ARMA models
6. Seasonal ARMA
7. Seasonal ARIMA models
Building ARIMA models

1. Plot the time series.
   Look for trends, seasonal components, step changes, outliers.
2. Nonlinearly transform data, if necessary
3. Identify preliminary values of $d$, $p$, and $q$.
4. Estimate parameters.
5. Use diagnostics to confirm residuals are white/iid/normal.
Identifying preliminary values of \( d \): Sample ACF

Trends lead to slowly decaying sample ACF:
Identifying preliminary values of $d$, $p$, and $q$

For identifying preliminary values of $d$, a time plot can also help.

Too little differencing: not stationary.
Too much differencing: extra dependence introduced.

For identifying $p$, $q$, look at sample ACF, PACF of $(1 - B)^d X_t$:

<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 $P, Q \geq 0$ and $s > 0$, we say that a time series $\{X_t\}$ is an \textbf{ARMA}(P,Q)$_s$ process if $\Phi(B^s)X_t = \Theta(B^s)W_t$, where

\[
\Phi(B^s) = 1 - \sum_{j=1}^{P} \Phi_j B^{js},
\]

\[
\Theta(B^s) = 1 + \sum_{j=1}^{Q} \Theta_j B^{js}.
\]

It is \textbf{causal} iff the roots of $\Phi(z^s)$ are outside the unit circle.
It is \textbf{invertible} iff the roots of $\Theta(z^s)$ are outside the unit circle.
Pure seasonal ARMA Models

Example: \( P = 0, Q = 1, s = 12. X_t = W_t + \Theta_1 W_{t-12} \).

\[
\begin{align*}
\gamma(0) &= (1 + \Theta_1^2) \sigma_w^2, \\
\gamma(12) &= \Theta_1 \sigma_w^2, \\
\gamma(h) &= 0 \quad \text{for } h = 1, 2, \ldots, 11, 13, 14, \ldots.
\end{align*}
\]

Example: \( P = 1, Q = 0, s = 12. X_t = \Phi_1 X_{t-12} + W_t. \)

\[
\begin{align*}
\gamma(0) &= \frac{\sigma_w^2}{1 - \Phi_1^2}, \\
\gamma(12i) &= \frac{\sigma_w^2 \Phi_1^i}{1 - \Phi_1^2}, \\
\gamma(h) &= 0 \quad \text{for other } h.
\end{align*}
\]
Pure seasonal ARMA Models

The ACF and PACF for a seasonal ARMA\((P,Q)_s\) are zero for \(h \neq si\). For \(h = si\), they are analogous to the patterns for ARMA\((p,q)\):

- **Model: AR**(P)s
  - ACF: decays
  - PACF: zero for \(i > P\)
- **Model: MA**(Q)s
  - ACF: zero for \(i > Q\)
  - PACF: decays
- **Model: ARMA**(P,Q)s
  - ACF: decays
  - PACF: decays
For $p, q, P, Q \geq 0$ and $s > 0$, we say that a time series $\{X_t\}$ is a multiplicative seasonal ARMA model $(\text{ARMA}(p,q) \times (P,Q)_s)$ if $\Phi(B^s) \phi(B) X_t = \Theta(B^s) \theta(B) W_t$.

If, in addition, $d, D > 0$, we define the multiplicative seasonal ARIMA model $(\text{ARIMA}(p,d,q) \times (P,D,Q)_s)$

$$\Phi(B^s) \phi(B) \nabla_s^D \nabla^d X_t = \Theta(B^s) \theta(B) W_t,$$

where the seasonal difference operator of order $D$ is defined by

$$\nabla_s^D X_t = (1 - B^s)^D X_t.$$
Multiplicative seasonal ARMA Models

Notice that these can all be represented by polynomials

$$\Phi(B^s)\phi(B) \nabla_s^D \nabla^d = \Xi(B), \quad \Theta(B^s)\theta(B) = \Lambda(B).$$

But the difference operators imply that $\Xi(B)X_t = \Lambda(B)W_t$ does not define a stationary ARMA process (the AR polynomial has roots on the unit circle). And representing $\Phi(B^s)\phi(B)$ and $\Theta(B^s)\theta(B)$ as arbitrary polynomials is not as compact.

How do we choose $p, q, P, Q, d, D$?

First difference sufficiently to get to stationarity. Then find suitable orders for ARMA or seasonal ARMA models for the differenced time series. The ACF and PACF is again a useful tool here.
Introduction to Time Series Analysis. Lecture 14.

1. Review: Maximum likelihood estimation
2. Computational simplifications: un/conditional least squares
3. Diagnostics
4. Model selection
5. Integrated ARMA models
6. Seasonal ARMA
7. Seasonal ARIMA models