Lecture 3, Estimation and model validation

Erik Lindström

FMS161/MASSM18 Financial Statistics
Maximum likelihood

Let \( x^{(N)} = (x_1, \ldots, x_N) \) be a sample from some parametric class of models with known density \( f_{x^{(N)}}(x_1, \ldots, x_n; \Theta) = L(x^{(N)}; \theta) \), where \( \theta \in \Theta \) is some unknown parameter vector.

The **Maximum Likelihood estimator** (MLE) is defined as

\[
\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \Theta} L(x^{(N)}; \theta)
\]  

Taking logarithm does not change the argument, so this is equivalently written as

\[
\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \Theta} \ell(x^{(N)}; \theta)
\]

where \( \ell(\theta) = \log L(x^{(N)}; \theta) \).
The asymptotic distribution for the MLE is given by

$$\sqrt{N} \left( \hat{\theta} - \theta \right) \overset{d}{\to} N \left( 0, I_N(\theta)^{-1} \right)$$

(3)

Theorem (Cramer-Rao)

Let $T(X_1, \ldots, X_N)$ be an unbiased estimator of $\theta$ then

$$V(T(X^N)) \geq I_N(\theta)^{-1} = - \left( E \left[ \nabla_\theta \nabla_\theta \log(L(x^{(N)}; \theta)) \right] \right)^{-1},$$

and the MLE attains this lower bound.
The asymptotic distribution for the MLE is given by

$$\sqrt{N} \left( \hat{\theta} - \theta \right) \xrightarrow{d} N \left( 0, I_N(\theta)^{-1} \right)$$

(3)

**Theorem (Cramer-Rao)**

Let $T(X_1, \ldots, X_N)$ be an unbiased estimator of $\theta$ then

$$\text{Var}(T(X^N)) \geq I_N(\theta)^{-1} = -\left( E \left[ \nabla_{\theta} \nabla_{\theta} \log(L(x^{(N)}; \theta)) \right] \right)^{-1},$$

and the MLE attains this lower bound.
Examination of the data

Before starting to do any estimation we should carefully look at the dataset.

- Are the data correct?
- Do the data contain outliers?
- Missing values?
- Do we have measurements of all relevant explanatory variables?
- Timing errors?
Model validation

There are two types of validation.

- **Absolute.** Are the model conditions fulfilled?
- **Relative.** Is the estimated model good enough, compared to some other model. Both can still be wrong...
There are two types of validation.

- Absolute. Are the model conditions fulfilled?
- Relative. Is the estimated model good enough, compared to some other model. Both can still be wrong...
Model validation

There are two types of validation.

▶ Absolute. Are the model conditions fulfilled?

▶ Relative. Is the estimated model good enough, compared to some other model. Both can still be wrong...
There are two types of validation.

- **Absolute.** Are the model conditions fulfilled?
- **Relative.** Is the estimated model good enough, compared to some other model. *Both can still be wrong...*
Model validation

There are two types of validation.

- **Absolute.** Are the model conditions fulfilled?
- **Relative.** Is the estimated model good enough, compared to some other model. Both can still be wrong...
We have some external knowledge of data e.g. underlying physics (Gray box models).

- Looking at the estimated parameters does the model make sense.
- Are effects going in the right directions?
- Do the parameters have reasonable values?
Residuals

The residuals $\{e\}$ Should be i.i.d. This implies:

No auto-dependence

$$\text{Cov}(f(e_n), g(e_{n+k})) = 0, \forall k \in \mathbb{Z}, \forall f, g,$$

such that $\mathbb{E}[f(e)^2] < \infty, \mathbb{E}[g(e)^2] < \infty$.

No cross correlation:

$$\text{Cov}(f(e_n), g(u_{n+k})) = 0, \forall k \in \mathbb{Z}, \forall f, g,$$

such that $\mathbb{E}[f(e)^2] < \infty, \mathbb{E}[g(u)^2] < \infty$ where $u$ is some external signal used as explanatory variable.
Residuals

The residuals \( \{e\} \) Should be i.i.d. This implies:

No auto-dependence

\[
\text{Cov}(f(e_n), g(e_{n+k})) = 0, \forall k \in \mathbb{Z}, \forall f, g,
\]

such that \( \mathbb{E}[f(e)^2] < \infty, \mathbb{E}[g(e)^2] < \infty \).

No cross correlation:

\[
\text{Cov}(f(e_n), g(u_{n+k})) = 0, \forall k \in \mathbb{Z}, \forall f, g,
\]

such that \( \mathbb{E}[f(e)^2] < \infty, \mathbb{E}[g(u)^2] < \infty \) where \( u \) is some external signal used as explanatory variable.
The residuals \( \{e\} \) should be i.i.d. This implies:

No auto-dependence

\[
\text{Cov}(f(e_n), g(e_{n+k})) = 0, \forall k \in \mathbb{Z}, \forall f, g,
\]

such that \( \mathbb{E}[f(e)^2] < \infty, \mathbb{E}[g(e)^2] < \infty \).

No cross correlation:

\[
\text{Cov}(f(e_n), g(u_{n+k})) = 0, \forall k \in \mathbb{Z}, \forall f, g,
\]

such that \( \mathbb{E}[f(e)^2] < \infty, \mathbb{E}[g(u)^2] < \infty \) where \( u \) is some external signal used as explanatory variable.
Normalized prediction errors

Residuals are usually normalized prediction errors

\[ e_n = \frac{y_n - \mathbb{E}[Y_n|\mathcal{F}_{n-1}]}{\sqrt{V(Y_n|\mathcal{F}_{n-1})}}. \]

This can in many cases also be generalized to SDE-models.
Formal tests

- Test for dependence in residuals (Box-Ljung).

\[ T = N(N+2) \sum_{k=1}^{l} \frac{\gamma(k)^2}{N-k}. \text{ Reject if } T > \chi^2_{1-\alpha,l}. \]

- Signtest on residuals \# of positive \( \in \text{Bin}(N, 1/2) \).

- Resimulate the model from residuals. Can it reproduce data?
Scatterplots of residuals

- $e_n$ vs $e_{n-1}$ (autocorr).
- $e_n$ vs $y_{n|n-1} = \mathbb{E}[y_n|\mathcal{F}_{n-1}]$ prediction error- remaining auto dependence.
- $e_n$ vs $u_n$ external dependence.
A good example (a well estimated AR(1) process)

\[ e_{n-1} \text{ vs } e_n \]

\[ e_n \text{ vs } y_{n|n-1} \]

**SACF**

**Normplot**
An example of wrong order (an AR(2) model estimated with a AR(1) model)

\[ e_{n-1} \text{ vs } e_n \]

\[ e_n \text{ vs } y_{n|n-1} \]

SACF

Normplot

Erik Lindström
Lecture 3, Estimation and model validation
An example of wrong model structure (a non-linear model estimated with a AR(1) model)

$$e_{n-1} \text{ vs } e_n$$

$$e_n \text{ vs } y_{n|n-1}$$

SACF

Normplot
Note that overfitting gives residuals that look good. Therefore it is important to test predictions also out of sample.

- Split data into an estimation and a validation set.
- Cross validation
Example overfitting (ARMA(1,1) fitted with ARMA(3,3))

\[ e_{n-1} \text{ vs } e_n \text{ insamp} \]

\[ e_{n-1} \text{ vs } e_n \text{ outsamp} \]

SACF insamp

SACF outsamp

Erik Lindström
Lecture 3, Estimation and model validation
Relative model validation

Test if a large model is really necessary.

\[ H_0 : \theta' = \theta'_0 \]
\[ H_1 : \theta' \text{ free}. \]

Hypothesis test Wald or LR.

Wald:

\[ l_{\hat{\theta}} = \hat{\theta} \pm \lambda_{\alpha/2} d(\hat{\theta}) \]
Let $Q(n)$ be the sum of squared residuals for an estimated model with $n$ parameters from $N$ observations.

Test $n_1$ vs $n_2$ parameters, then for true order $n_0 \leq n_1 < n_2$

i) $\frac{Q(n_2)}{\sigma^2} \in \chi^2(N - n_2)$.

ii) $\frac{Q(n_1) - Q(n_2)}{\sigma^2} \in \chi^2(n_2 - n_1)$.

iii) $Q(n_2)$ and $Q(n_1) - Q(n_2)$ are independent.

iv) $\eta = \frac{N-n_2}{n_2-n_1} \frac{Q(n_1) - Q(n_2)}{Q(n_2)} \in F(n_2 - n_1, N - n_2)$.

If $\eta$ is large pick model 2 else pick model 1. This is an exact test for AR models.
Asymptotic tests

\[ LR = -2 \left( \log(L(\theta^{\text{Model}1}) - \log(L(\theta^{\text{Model}2})) \right) \]

If model 1 has \( n_1 \) parameters and model 2 has \( n_2 \) parameters \( n_2 > n_1 \) then \( LR \) is asymptotically distributed as

\[ LR \overset{d}{\to} \chi^2(n_2 - n_1). \]  

(4)

This is true for all models where the likelihood regularity conditions apply (a very large class of distributions) if \( N \) is large. This is the most powerful test in the sense of Neyman-Pearson.
The main idea is to penalize too many parameters.

- AIC (Akaike's Information Criteria): $-2 \log(L(\theta)) + 2 \dim(\theta)$, where the minimum is taken as $\dim(\theta)$. Problem it overestimates the model order.

- BIC (Bayesian Information Criteria): $-2 \log(L(\theta)) + 2 \dim(\theta) \log(N)$. This slightly underestimates the model order.

- Alternative LIL (law of iterated logarithm): $-2 \log(L(\theta)) + 2 \dim(\theta) \log(\log(N))$. 
Example choice of model AR(3) process

The number of observations is 500 the number of replicates is 200

<table>
<thead>
<tr>
<th>crit/order</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>≥11</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>13</td>
<td>11</td>
<td>15</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>89</td>
</tr>
<tr>
<td>BIC</td>
<td>0</td>
<td>78</td>
<td>122</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LIL</td>
<td>0</td>
<td>4</td>
<td>140</td>
<td>29</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>