Forecasting and Estimation

Helle Bunzel

ISU

February 3, 2009
Very frequently the goal of estimating time series is to provide forecasts of future values.

This typically means you treat the data differently than if you were simply fitting a model.

A crucial step in choosing your forecast is to define a loss function. Examples are:

- Minimizing the absolute difference between the forecast and the eventual value.
- Minimizing the square difference.
- Asymmetric loss functions which penalize too high forecasts more than too low (GDP, state budgets).

This is a potential topic for end-of-semester presentation.

Many other options for loss functions.
Forecasting II

- It is typically not the case that the model that provides the best fit is the best model for forecasting.
- Example: Consider the following ARMA(2,1) model:

\[
y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta \varepsilon_{t-1}
\]

- If we knew the true model, we would forecast \( y_{t+1} \) at time \( t \) as

\[
y_{t+1|t} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + \theta \varepsilon_t
\]

- We’d make the mistake of:

\[
y_{t+1} - y_{t+1|t} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + \varepsilon_{t+1} + \theta \varepsilon_t
\]

\[
- (\mu + \phi_1 y_t + \phi_2 y_{t-1} + \theta \varepsilon_t)
\]

\[
= \varepsilon_{t+1}
\]
Forecasting III

- Only the inherently unpredictable white noise error.
- In reality, however, we have to use

\[ y_{t+1|t}^* = \hat{\mu} + \hat{\phi}_1 y_t + \hat{\phi}_2 y_{t-1} + \hat{\theta} \hat{\epsilon}_t \]

- The mistake we make is:

\[
y_{t+1} - y_{t+1|t} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + \epsilon_{t+1} + \theta \epsilon_t \\
- (\hat{\mu} + \hat{\phi}_1 y_t + \hat{\phi}_2 y_{t-1} + \hat{\theta} \hat{\epsilon}_t) \\
= \epsilon_{t+1} + (\mu - \hat{\mu}) + (\phi_1 - \hat{\phi}_1) y_t \\
+ (\phi_2 - \hat{\phi}_2) y_{t-1} + (\theta - \hat{\theta}) \epsilon_t \\
+ (\epsilon_t - \hat{\epsilon}_t) \hat{\theta}
\]
Clearly the estimation of the parameters and the residuals add to the error we make.

- We cannot do much about the residuals.
- It is possible that approximating the model by an AR(1) or an MA(1) may lead to more accurate forecast.

For goodness of fit, this does not happen. There the "true" model always gives us the best result.
Forecasts with MSE loss I

- More generally we might have more regressors than just lagged values of $y$.
  - Let $X_t$ be the regressors, possibly containing lagged values of $y$.
- The most commonly used loss function would be the *quadratic* loss function or the *Mean Squared Error* associated with the forecast:

  $$MSE\left(y_{t+1|t}\right) = E\left(y_{t+1} - y^*_{t+1|t}\right)^2$$

- Once a loss function is determined, the forecast is chosen to minimize this function.

  **Result** The forecast which minimizes the $MSE$ loss function is

  $$y^*_{t+1|t} = E\left(y_{t+1|X_t}\right)$$
Forecasts with MSE loss II

- In general this is as much as we can say.
- We can make more progress if we restrict our attention to linear forecasts. That is:

\[ y_{t+1|t} = \alpha' X_t \]

**Result** The *linear* forecast which minimizes the *MSE* loss among all possible *linear forecasts* is the projection of \( y_{t+1} \) onto \( X_t \):

\[
\alpha' = E[y_{t+1}X_t] E \left[ \left( X'_t X_t \right)^{-1} \right] \
\alpha = E \left[ \left( X'_t X_t \right)^{-1} \right] E \left[ X'_t y_{t+1} \right]
\]
Forecasts with MSE loss III

and

\[ y_{t+1|t}^* = \alpha' X_t = E \left[ y_{t+1} \right] E \left[ \left( X_t' X_t \right)^{-1} \right] X_t \]

Note that this expression is closely connected to the OLS estimator from the regression:

\[ y_{t+1} = \beta X_t + \varepsilon_t \]

where

\[ \hat{\beta} = \left( \frac{1}{T} \sum_{t=1}^{T} X_t' X_t \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} X_t' y_{t+1} \]
Forecasts with MSE loss IV

If the process \( \{ y_{t+1}, X_t \} \) is covariance stationary and ergodic for second moments, then as \( T \to \infty \)

\[
\frac{1}{T} \sum_{t=1}^{T} X_t' X_t \xrightarrow{P} E(X_t' X_t)
\]

and

\[
\frac{1}{T} \sum_{t=1}^{T} X_t' y_{t+1} \xrightarrow{P} E(X_t' y_{t+1})
\]
If this holds, we have

$$\hat{\beta}^P \rightarrow \alpha$$

and OLS can be used to find the optimal linear forecasts.

Pay attention to violations of stationarity!

Note that we need less assumptions than standard OLS. We are simply detecting patterns.
Forecasts Based on an Infinite Number of Known Errors

- For now assume that we have an infinite number of observations.
- First consider a process with an $MA(\infty)$ representation such that:

\[ y_t = \mu + \psi(L) \varepsilon_t \]

where

\[ \psi(L) = \sum_{j=0}^{\infty} \psi_j L^j, \quad \psi_0 = 1, \]

\( \{ \varepsilon_t \} \) is white noise and

\[ \sum_{j=0}^{\infty} |\psi_j| < \infty \]
Forecasts Based on an Infinite Number of Known Errors II

Suppose that we know the process up to and including date $t$ and want to forecast $y_{t+s}$.

Note that

$$y_{t+s} = \mu + \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \ldots + \psi_{s-1} \epsilon_{t+1} + \psi_s \epsilon_t + \ldots.$$ 

The optimal linear forecast is the projection of $y_{t+s}$ on all known $\epsilon_t$ and a constant. Hamilton denotes this:

$$y_{t+s|t}^* = \hat{E}[y_{t+s}|\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots] = \mu + \psi_s \epsilon_t + \psi_{s+1} \epsilon_{t-1} + \ldots.$$ 

The forecast error is:

$$y_{t+s} - y_{t+s|t}^* = \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \ldots + \psi_{s-1} \epsilon_{t+1}$$
Forecasts Based on an Infinite Number of Known Errors III

And the mean squared error is:

\[ E \left[ (y_{t+s} - y_{t+s}^*)^2 \right] = \sigma^2 (1 + \psi_1^2 + \ldots + \psi_{s-1}^2) \]

If we instead of a \( MA(\infty) \) process had a \( MA(q) \) process, the same principle can be applied.

We get that, if \( q \geq s \)

\[ y_{t+s} = \mu + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \ldots + \psi_{s-1} \varepsilon_{t+1} + \psi_s \varepsilon_t + \ldots + \psi_q \varepsilon_{t+s-q} \]

and if \( q < s \)

\[ y_{t+s} = \mu + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \ldots + \psi_q \varepsilon_{t+s-q} \]

where

\[ t + s - q > t \]
Forecasts Based on an Infinite Number of Known Errors IV

- The optimal linear forecast is:
  - For $q \geq s$
    $$y_{t+s|t}^* = \hat{E} \left[ y_{t+s} \mid \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \right] = \mu + \psi_s \varepsilon_t + \ldots + \psi_q \varepsilon_{t-q}$$
  - For $q < s$
    $$y_{t+s|t}^* = \hat{E} \left[ y_{t+s} \mid \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \right] = \mu$$

- The MSE is:
  - For $q \geq s$
    $$y_{t+s} - y_{t+s|t}^* = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \ldots + \psi_{s-1} \varepsilon_{t+1}$$

  and

  $$E \left[ \left( y_{t+s} - y_{t+s|t}^* \right)^2 \right] = \sigma^2 \left( 1 + \psi_1^2 + \ldots + \psi_{s-1}^2 \right)$$
Forecasts Based on an Infinite Number of Known Errors V

For \( q < s \)

\[ y_{t+s} - y^*_t = y_{t+s} - \mu = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \ldots + \psi_q \varepsilon_{t+s-q} \]

and

\[ E \left[ (y_{t+s} - y^*_t)^2 \right] = \sigma^2 \left( 1 + \psi_1^2 + \ldots + \psi_q^2 \right) \]

- A complicated but useful way of writing this.
- Consider the polynomial

\[ \psi(L) = \sum_{j=0}^{q} \psi_j L^j, \psi_0 = 1 \]

- \( q \) can be finite or infinite.
Forecasts Based on an Infinite Number of Known Errors VI

- Now divide by $L^s$:

$$\frac{\psi(L)}{L^s} = \sum_{j=0}^{q} \psi_j L^{j-s}, \psi_0 = 1$$

- The *Annihilation Operator* replaces all $L^j$ with $j < 0$ with 0. The notation is:

$$\left[ \frac{\psi(L)}{L^s} \right]_+ = \sum_{j=s}^{q} \psi_j L^{j-s}$$

- This one makes finding the forecast easier.

- Recall that the forecast projects onto the errors starting $s$ periods earlier.
Forecasts Based on an Infinite Number of Known Errors VII

- For the $MA(\infty)$ process we had:

$$y_{t+s|t}^* = \mu + \psi_s \epsilon_t + \psi_{s+1} \epsilon_{t-1} + \ldots.$$ 

- This can easily be written as

$$y_{t+s|t}^* = \mu + \left[ \frac{\psi(L)}{L^s} \right] \epsilon_t = \mu + \sum_{j=s}^{\infty} \psi_j L^{j-s} \epsilon_t$$

$$= \mu + \sum_{j=s}^{\infty} \psi_j \epsilon_{t+s-j} = \mu + \sum_{i=0}^{\infty} \psi_{s+i} \epsilon_{t-i}$$

- So, this prediction formula is based on an $MA(q)$ representation (with $q \leq \infty$) and an infinite number of errors.
Forecasts Based on an Infinite Number of Observations I

- While we can calculate all these theoretically, we wish to forecast using the data, that is past values of $y$, not the errors.
- Suppose we have a well-behaved process. Basically:
  - AR process is stable and stationary
  - MA process is invertible (details in next problemset)
  - AR coefficients are absolutely summable.
- If these are true, we can use the observations $\{y_i\}_{i=-\infty}^{t}$ to find the errors $\{\varepsilon_i\}_{i=-\infty}^{t}$.
Forecasts Based on an Infinite Number of Observations II

Examples:

Suppose we have an AR(1) process:

\[ y_t = c + \phi y_{t-1} + \epsilon_t \]

then we can find

\[ \epsilon_i = (1 - \phi L) y_i - c \]

or

\[ \epsilon_i = (1 - \phi L) (y_i - \mu) \]
Forecasts Based on an Infinite Number of Observations III

Suppose we have an $MA(1)$ process:

$$y_t = \mu + (1 + \theta L) \varepsilon_t$$

Now, if $|\theta| < 1$

$$(1 + \theta L)^{-1} (y_t - \mu) = \varepsilon_t$$

So

$$\varepsilon_t = \sum_{i=0}^{\infty} \theta^i L^i (y_t - \mu)$$
Forecasts Based on an Infinite Number of Observations IV

- The optimal linear forecast is:

  - For $s = 1$
    \[
    y_{t+s | t}^* = \hat{E} [y_{t+s | t} \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots] = \mu + \theta \sum_{i=0}^{\infty} \theta^i L^i (y_t - \mu)
    \]

  - For $1 < s$
    \[
    y_{t+s | t}^* = \hat{E} [y_{t+s | t} \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots] = \mu
    \]

- Using the annihilation operator, we would have:

  \[
  y_{t+s | t}^* = \mu + \left[ \frac{\psi(L)}{L^s} \right] \epsilon_t
  \]
Forecasts Based on an Infinite Number of Observations

For $s = 1$

$$y^*_{t+1|t} = \mu + \left[ \frac{\psi(L)}{L} \right]_+ \epsilon_t = \mu + \theta \epsilon_t$$

$$= \mu + \theta \sum_{i=0}^{\infty} \theta^i L^i (y_t - \mu)$$

For $1 < s$

$$y^*_{t+s|t} = \mu + \left[ \frac{1 + \theta L}{L^s} \right]_+ \epsilon_t = \mu$$
Forecasts Based on an Infinite Number of Observations

- Under the assumptions we made earlier any standard processes can be written as $AR(\infty)$ processes.

- We can the find the $\varepsilon_t$ as

$$\eta(L)(y_t - \mu) = \varepsilon_t$$

where

$$\eta(L) = \sum_{j=0}^{\infty} \eta_j L^j$$
Forecasts Based on an Infinite Number of Observations VII

Then we can write

\[ y_{t+s|t} = \mu + \left[ \frac{\psi(L)}{L_s} \right] \epsilon_t \]

where \( \psi(L) = \eta(L)^{-1} \).

To get this as a function of the data, as opposed to the errors, we then write:

\[
y_{t+s|t}^* = \mu + \left[ \frac{\psi(L)}{L_s} \right] \eta(L)(y_t - \mu) + \psi(L)^{-1}(y_t - \mu)
\]

This is called the \textit{Wiener-Kolmogorov prediction formula}. 