Forecast uncertainty and Forecast Intervals

measure of the uncertainty of that estimate
One measure of the coefficients of a forecast is its root mean square forecast error (RMSFE).
The forecast error consists of two components
1. uncertainty arising from estimation of the regression coefficients.
2. uncertainty associated with the future unknown value of \( u_t \).

few coefficients and many observations ⇒ 2 > 1
ADL(1,1) model

\[
Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 x_{t-1} + u_t,
\]

where \( u_t \) is homoskedastic.

\[
\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\delta}_1 x_T
\]

\[
Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) Y_T + (\hat{\delta}_1 - \delta_1)x_T]
\]

\[
MSFE = E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]
\]

\[
= \sigma_u^2 + var[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) Y_T + (\hat{\delta}_1 - \delta_1)x_T],
\]

\[
RMSFE = \sqrt{MSFE}
\]

Forecast intervals

a 95% forecast interval
Assume that \( u_{T+1} \) is normally distributed,
The forecast error is normally distributed with variance equaling the MSFE.
→ a 95% C.I.

\[
\hat{Y}_{T+1|T} \pm 1.96 * SE(Y_{T+1} - \hat{Y}_{T+1|T})
\]

Lag Length selection using Information Criteria

How many lags should be included in a time series regression
put little lag variables → lose lots of information
put many lag variables → forecast error ↑

1. The F-statistic approach To choose \( p \) is start with a model with many lags and
to perform hypothesis test on the final lag,  
Example: AR(6)  
$H_0 : \beta_6 = 0 \rightarrow \text{significant}$, If not drop it.  
AR(5)

The drawback: it will produce too large a model.

2. The BIC
To estimate $p$ by minimizing an "information criterion"
Bayes information criterion (BIC) or
Schwarz information criterion (SIC)

$$BIC(p) = \ln \left(\frac{SSR(p)}{T}\right) + (p + 1)\frac{\ln T}{T}$$

Where $SSR(p)$ is the sum of squared residuals of the estimate AR(p).
The BIC estimator of $p$, $\hat{p}$ is the value that minimizes $BIC(p)$ among the possible choices $p = 0, 1, 2, \ldots, p_{\text{max}}$.
The first term: add one more lag, $RSS(SSR) \downarrow$,
The second term: add one more lag, $(p + 1)\frac{\ln T}{T} \uparrow$

The BIC trades off these two forces so that the number of lags that minimizes the BIC is a consistent estimator of the true lag length.

3. the AIC
Akaike information criterion (AIC)

$$AIC(p) = \ln \left(\frac{SSR(p)}{T}\right) + (P + 1)\frac{2}{T}$$

Note: $\ln T$ in the BIC, 2 in the AIC
The second term is the AIC is smaller
In large samples the AIC will overestimate $p$ with nonzero probability.
The BIC might yield a model with too few lags.
The AIC might provides a reasonable alternative.
Trends Nonstationarity

what is a trend?
(1) A trend is a persistent long-term movement of a variable over time.
(2) Deterministic v.s stochastic trends
There are two types of trends: deterministic/stochastic

A deterministic trend is a nonrandom function of time
e.g. linear in time
A stochastic trend is random and varies over time
The treatment of trends in economic time series focuses on stochastic rather than deterministic trends.

The random walk model of a trend

A time series $Y_t$ is a random walk if the change in $Y_t$ is iid, if

$$Y_t = Y_{t-1} + u_t,$$

where $u_t$ is iid.

$$E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$$

The basic idea of random walk is that the value of the series tomorrow is its value today, plus an unpredictable change.

$$E(Y_t|Y_{t-1}, Y_{t-2}, ...) = Y_{t-1},$$

If $Y_t$ follows a random walk, then the best forecast of tomorrow’s value is its value today.

A random walk with drift

$$Y_t = \beta_0 + Y_{t-1} + u_t,$$

where $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$. $\beta_0$ is the ”drift” in the random walk. If $\beta_0 > 0, Y_t \uparrow$
The best forecast of the series tomorrow is the value of the series today, plus the drift
\[ \beta_0. \]

**A random walk is nonstationary? Yes**

The variance of a random walk increases over time.

\[ \text{var}(Y_t) = \text{var}(Y_{t-1}) + \text{var}(u_t), \]

1. \( \text{var}(Y_t) = \text{var}(Y_{t-1}) \) if \( \text{var}(u_t) = 0 \),
2. \( Y_t = 0, Y_1 = u_1, \)
   \( Y_2 = u_1 + u_2, \)
   \( Y_t = u_1 + u_2 + \ldots + u_t, \)
Because \( u_t \) is serially uncorrelated,

\[ \text{var}(Y_t) = t \cdot \sigma_u^2 \]

the variance of \( Y_t \) depends on \( t \).
\( t \uparrow, \text{var}(Y_t) \uparrow \)
it is nonstationary.
The population autocorrelations are not defined.
The sample autocorrelations tend to be very close to one the \( j \)th sample autocorrelation of a random walk converges to one in probability.