Characterizing Forecast Uncertainty – Prediction Intervals

The estimated AR (and VAR) models generate *point forecasts* of $y_{t+s}$, $\hat{y}_{t+s,t}$.

Under our assumptions the point forecasts are asymptotically unbiased but biased in finite samples. (What does this mean? Why biased?)

We understand that the actual value of $y_{t+s}$ will almost certainly be greater or less than the point forecast. How can we characterize the amount of uncertainty regarding the size of the forecast error? *Prediction Intervals*
A 95% prediction interval for $y_{t+s}$ will be an interval of the form

$$\hat{y}_{t+s, t} + c$$

where $c$ is chosen so that

$$\Pr \{ \hat{y}_{t+s} - c \leq y_{t+s} \leq \hat{y}_{t+s} + c \} = 0.95$$

That is, 95-percent of the time the interval constructed according to this procedure will contain the true $y_{t+s}$. 90%, 99%, 80% intervals are defined similarly (i.e., by choosing $c$ so that $\Pr \{ . \} = \ldots$).

How to choose $c$?
There are two sources of forecast errors:

• fundamental uncertainty: future ε’s
• sampling error: using estimated a’s

For example, consider the AR(1) model –

\[ y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1} \]
\[ \hat{y}_{t+1,t} = \hat{a}_0 + \hat{a}_1 y_t \]

and, the one-step-ahead forecast error is

\[ y_t - \hat{y}_{t+1,t} = (a_0 - \hat{a}_0) + (a_1 - \hat{a}_1) y_t + \varepsilon_{t+1} \]

Under appropriate side conditions on the ε’s the asymptotically valid 95-percent prediction interval for \( y_{t+1} \) will be

\[ \hat{y}_{t+1,t} \pm 1.96 \times s.d.(y_{t+1} - \hat{y}_{t+1,t}) \]

where s.d. (.) denotes the standard deviation of (.)
Often, forecast intervals are constructed ignoring sampling error, because these intervals are easier to construct. While easier to compute, these intervals will tend to underestimate the amount of forecast uncertainty.

In the AR(1) example, ignoring sampling error, the one-step ahead forecast error is $\varepsilon_{t+1}$ and the 95-percent PI for $y_{t+1}$ will be

$$\hat{y}_{t+1,t} \pm 1.96 \times \sigma_\varepsilon$$

We can replace $\sigma_\varepsilon$ with any consistent estimator (e.g., the OLS or MLE estimator) without affecting the asymptotic validity of this interval.
What are the 95-percent PI for the AR(1) s-step ahead forecasts (if we ignore uncertainty due to sampling error)?

What are the 95-percent PI for AR(p) s-step ahead forecasts (if we ignore uncertainty due to sampling error)?
Accounting for the effects of sampling errors on forecast uncertainty becomes quite cumbersome computationally, particularly as \( p \) increases, if we rely on theory. (For example, in the AR(1) case, it will depend not only on a second-moments, but also depend on the third and fourth moments.)

One practically and theoretically appealing approach to constructing prediction intervals that account for fundamental and parameter uncertainty and that can be applied in a wide variety of settings and conditions is the *bootstrap prediction interval*.

We will provide an introduction to bootstrap prediction intervals through a couple of examples.
Digression on Bootstrap Procedures –
Bootstrap procedures provide an alternative approach to exact or asymptotic distribution theory to do statistical inference (interval estimation; hypothesis testing).

These procedures are particularly useful when standard distribution theory does not provide useful or convenient results or may require conditions that seem inappropriate in a particular setting or are difficult to work with analytically.

Bootstrap methods can be applied in a wide variety of settings and can be shown (under the right conditions) to provide the same distributional results as asymptotic theory and, in many settings convergence to the limiting distribution occurs more rapidly with the bootstrap.

The cost? Computationally expensive. Not so much of an issue anymore with current desktop and laptop technology.
Simple example of a bootstrap application –

Suppose we draw \( y_1, \ldots, y_n \) randomly with replacement from a population, \( Y \).

We are interested in estimating the population mean of \( Y \), \( \mu \). An unbiased (and consistent) estimator of \( \mu \) is the sample mean, \( \hat{\mu} = \frac{1}{n} \sum_{i} y_i \).

In order to construct a confidence interval for \( \mu \) or test a hypothesis about \( \mu \), we need to know or estimate the distribution of \( \hat{\mu} \). If we could resample from \( Y \) as often as we would like, this would be straightforward. The problem is that we are usually not in a position to resample.

If we assume normality then we have an exact result for this distribution. Otherwise, we could apply the CLT to get the asymptotic distribution.
Approximating the distribution of $\hat{\mu}$ using a bootstrap procedure –

**Act as though $Y=\{y_1,\ldots,y_n\}$.**

Let $Y^* = \{y_1,\ldots,y_n\}$

1. Draw $n$ times randomly with replacement from $Y^*$ to get the bootstrap sample, $(y_1^{(b)},\ldots,y_n^{(b)})$.

2. Use the sample mean mean of this bootstrap sample to get a bootstrap estimate of $\mu$, $\hat{\mu}^{(b)} = \frac{1}{n} \sum_{i=1}^{n} y_i^{(b)}$.

3. Redo steps 1 and 2 a large number, $B$, times to get $\hat{\mu}^{(b)}$, $b = 1,\ldots,B$. Use the observed distribution of the $\hat{\mu}^{(b)}$’s as an estimator of the distribution of $\hat{\mu}$.

A nice introduction to bootstrap methods:


End Digression
Bootstrapping the prediction interval for the AR(1) model –

Suppose that a stationary time series $y_t$ is assumed (possibly after data-based pre-testing) to follow the AR(1) model

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$$

where the $\varepsilon$’s are an i.i.d. w.n. series.

We have a data set $y_0, y_1, \ldots, y_T$ and we want to forecast $y_{T+s}, s = 1, \ldots, H$: point forecasts and 95-percent forecast intervals.
1. Fit the model to the data by OLS to get
\( \hat{\alpha}_0, \hat{\alpha}_1, \hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T \)

2. Construct the s-step ahead point forecasts

\[
\hat{y}_{T+s} = (1 + \alpha_1 + \ldots + \alpha_1^{s-1})\alpha_0 + \alpha_1^s y_T
\]

for \( s = 1, 2, \ldots, H \)

3. Construct bootstrap 95-percent forecast intervals.
   
   i. randomly draw with replacement
   \( \hat{\varepsilon}_1^{(b)}, \ldots, \hat{\varepsilon}_T^{(b)} \) from \( \{ \hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T \} \) [How?]
   
   ii. construct the bootstrap sample
   \( y_1^{(b)}, \ldots, y_T^{(b)} \) according to

\[
y_t^{(b)} = \hat{\alpha}_0 + \hat{\alpha}_1 y_{t-1}^{(b)} + \hat{\varepsilon}_t^{(b)}
\]

for \( t = 1, \ldots, T \) where we can either fix \( y_0 \) at its actual value or draw \( y_0^{(b)} \) randomly from \( \{ y_0, \ldots, y_T \} \).

Note that the \( \alpha \)-hats are the OLS estimates from Step 1.
iii. Fit $y_1^{(b)}, \ldots, y_T^{(b)}$ to an AR(1) model (by OLS) to get estimates of $\alpha_0$ and $\alpha_1$, $\hat{\alpha}_0^{(b)}$ and $\hat{\alpha}_1^{(b)}$, then use these to generate the s-step ahead forecasts

$$
\hat{y}_{T+s}^{(b)} = (1 + \hat{\alpha}_1^{(b)} + \ldots + \hat{\alpha}_1^{(b)s-1})\hat{\alpha}_0^{(b)} + \hat{\alpha}_1^{(b)s} y_T^{(b)}
$$

for $s = 1, \ldots, H$.

iv. Do (i)-(iii) a “large number” of time: $B$.

v. Use the frequency distribution of $\hat{y}_{T+s}^{(b)}$, $b = 1, \ldots, B$ to construct the 95-percent forecast interval for $y_{T+s}$.

For example, if $B = 10,000$ then the 95-percent FI for $y_{T+s}$ would be the interval from the 250-th to 9750-th ordered values of $\hat{y}_{T+s}^{(b)}$. 
How would you generalize this procedure to construct forecast intervals for AR(p) or VAR(p) models?