- Forecasting means using the nonlinear model estimated on samples $\{x_n\}$, $n = 1, \ldots, N$, to perform out-of-sample (true) prediction.
- Questions arising are:
  - What is the quality of the prediction performed?
  - What happens when multi-step ahead prediction is to be made?

- We have already seen that the optimal predictor of x_{n+h} at time n in the least-squares sense is:

$$\hat{x}_{n+h|n} = \mathsf{E}[\hat{x}_{n+h} \mid \Omega_n]$$

where Ω_n sums up all the information available up to time n. In the nonlinear case, it is generally impossible to know in advance whether it is better to directly estimate this predictor or to iterate the 1-step ahead one.

- For a linear AR model, multi-step forecasting is easy. If the model is:

$$x_n = a_1 x_{n\text{-}1} + \ldots + a_p x_{n\text{-}p} + \varepsilon_n$$

Then the 1-step ahead forecast is:

$$\hat{x}_{n+1|n} = a_1 x_n + \cdots + a_1 x_{n-p+1}$$

And the forecast error is ε_{n+1}. The 2-step ahead forecast is:

$$\hat{x}_{n+2|n} = a_1 \hat{x}_{n+1|n} + \cdots + a_1 x_{n-p+2}$$

- And so on. It can be shown by recursion that the mean-square prediction error MSPE(h) for h-step ahead prediction is given by:

$$\mathrm{MSPE}(h) = \sigma^2 \sum_{i=0}^{h-1} b_i^2$$

With σ^2 the variance of $\{\varepsilon_n\}$, and $\{c_i\}$ the impulse response of the AR filter. Basically, all goes well because the linear operator and the expectation commute.

- When the model is NAR:

$$x_n = g(\boldsymbol{x}_{n-1}) + \varepsilon_n$$

the optimal 1-step ahead forecast is:

$$\hat{x}_{n+1|n} = \mathrm{E}[x_{n+1} \mid \Omega_n] = g(\boldsymbol{x}_n)$$

- But the optimal 2-step ahead predictor is:

$$\hat{x}_{n+2|n} = \mathrm{E}[x_{n+2} \mid \Omega_n] = \mathrm{E}[g(\boldsymbol{x}_{n+1}) \mid \Omega_n]$$

and alas:

$$\mathrm{E}[g(\boldsymbol{x}_{n+1}) \mid \Omega_n] \neq g(\mathrm{E}[x_{n+1} \mid \Omega_n]) = g(\hat{x}_{n+1|n})$$

- This means that iterating the NAR function to produce successive forecasts is not a good idea. It can be shown to produce biased values.
- The right relation between 1- and 2-step ahead prediction is:

$$\hat{x}_{n+2|n} = \mathsf{E}[g(g(\boldsymbol{x}_n) + \varepsilon_{n+1}) \mid \Omega_n]$$
$$= \mathsf{E}[g(\hat{x}_{n+1|n} + \varepsilon_{n+1}) \mid \Omega_n]$$

- The problem comes of course from the term ε_{n+1}. A possible solution is to compute:

$$\hat{x}_{n+2|n} = \int_{-\infty}^{\infty} g(\hat{x}_{n+1|n} + \varepsilon) f(\varepsilon) d\varepsilon$$

where *f*(.) is the probability density of the innovations $\{\varepsilon_n\}$. Two problems: *f*(.) is generally only imperfectly known, and this integral may be hard to compute analytically. One resorts to numerical methods.

- An approximate way to compute this forecast is using a Monte Carlo approach:

$$\hat{x}^{(\mathrm{mc})}_{n+2|n} = \frac{1}{K}\sum_{k=1}^{K} g(\hat{x}_{n+1|n} + \nu_k)$$

With the random variables $\{\nu_k\}$ drawn having the presumed probability density $f(.)$.

- The difficulty of course is the state of knowledge about $f(.)$.

- A solution to this is to use a *bootstrap* approach, i.e. build an estimate:

$$\hat{x}^{(b)}_{n+2|n} = \frac{1}{K}\sum_{k=1}^{K} g(\hat{x}_{n+1|n} + \hat{\varepsilon}_k)$$

where the $\{e_k\}$ are innovations $\{\varepsilon_n\}$ obtained with the model on the N available samples, and are drawn from this set of innovations at random with replacement. The bootstrap is a bit poorer in performance, but no assumption on innovation distribution has to be made.

- It is to be noted that both Monte Carlo and bootstrap approaches allow one to compute *interval* forecasts.
- Indeed, instead of computing only the average as shown before, it is possible to estimate the probability density of the forecasts.
- This is particularly appealing since for nonlinear models this density can be asymmetrical and even multimodal.

1. P. H. Franses and D. van Dijk, *Non-Linear Time Series Models in Empirical Finance*, Cambridge Univ. Press, 2000.