Multistep-ahead conformal prediction

1 Properties of multi-step ahead forecast errors

Refer to Harvey, Leybourne, and Newbold (1997), Diebold (2017), and Sommer (2023).

1.1 Key properties of optimal forecasts

- P1: Optimal forecasts are unbiased.
- P2: Optimal forecasts have 1-step-ahead errors that are white noise.
- **P3: Optimal forecasts have h-step-ahead errors that follows an approximate MA$(h-1)$ process.**

1.2 Proof of P3

Assuming that a univariate time series $y_1, \cdots, y_T$ is generated by the non-stationary autoregressive process

$$y_t = f_{\boldsymbol{\theta}_t}(\boldsymbol{x}_{t-1}) + \epsilon_t,$$

where $\boldsymbol{x}_{t-1} = (y_{t-1}, \cdots, y_{t-d})'$, f is assumed to be a non-linear function in the vector $\boldsymbol{x}_{t-1}$ of lagged endogenous variable, and innovations $\{\epsilon_t\}$ are assumed to be white noise.

- For 2-step ahead forecast error

$$
\begin{aligned}
y_{t+2} &= f_{\boldsymbol{\theta}_{t+2}}(\boldsymbol{x}_{t+1}) + \epsilon_{t+2} \\
&= f_{\boldsymbol{\theta}_{t+2}}(y_{t+1}, \dots, y_{t-d+2}) + \epsilon_{t+2} \\
&= f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t) + \epsilon_{t+1}, y_t, \dots, y_{t-d+2}) + \epsilon_{t+2} \\
&\underset{te}{\approx} f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}) + \epsilon_{t+1} \frac{\partial f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2})}{\partial x_1} + \epsilon_{t+2} \\
&= f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}) + e_{t+2|t},
\end{aligned}
$$

where te means the first order Taylor series expansion of the function $f_{\boldsymbol{\theta}_{t+2}}$ at the point $(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2})$. The expansion shows that the sequence $\{e_{t+2|t}\}$ is at most serially correlated up to lag 1, which also indicates that $\{e_{t+2|t}\}$ follows a MA(1) process.

- For 3-step ahead forecast error

$$
\begin{aligned}
y_{t+3} =& f_{\boldsymbol{\theta}_{t+3}}(\boldsymbol{x}_{t+2}) + \epsilon_{t+3} \\
=& f_{\boldsymbol{\theta}_{t+3}}(y_{t+2}, y_{t+1}, \dots, y_{t-d+3}) + \epsilon_{t+3} \\
=& f_{\boldsymbol{\theta}_{t+3}}(f_{\boldsymbol{\theta}_{t+2}}(\boldsymbol{x}_{t+1}) + \epsilon_{t+2}, f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t) + \epsilon_{t+1}, y_t, \dots, y_{t-d+3}) + \epsilon_{t+3} \\
=& f_{\boldsymbol{\theta}_{t+3}}(f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}) + e_{t+2|t}, f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t) + \epsilon_{t+1}, y_t, \dots, y_{t-d+3}) + \epsilon_{t+3} \\
\underset{te}{\approx}& f_{\boldsymbol{\theta}_{t+3}}(f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}), f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+3}) \\
&+ e_{t+2|t} \frac{\partial f_{\boldsymbol{\theta}_{t+3}}(f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}), f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+3})}{\partial x_1} \\
&+ \epsilon_{t+1} \frac{\partial f_{\boldsymbol{\theta}_{t+3}}(f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}), f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+3})}{\partial x_2} + \epsilon_{t+3} \\
=& f_{\boldsymbol{\theta}_{t+3}}(f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+2}), f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \dots, y_{t-d+3}) + e_{t+3|t}
\end{aligned}
$$

So $e_{t+3|t}$ is a function of $\epsilon_{t+1}, \epsilon_{t+2}, \epsilon_{t+3}$, as $e_{t+2|t}$ is dependent on ϵ_{t+1} and ϵ_{t+2}.

- For h-step ahead forecast error, ..., $\{e_{t+h|t}\}$ follows a MA$(h-1)$ process.

1.3 Relationship between the h-step ahead forecast error and past errors

First write

$$
y_{t+h} = \hat{y}_{t+h|t} + e_{t+h|t},
$$

where

- $\hat{y}_{t+1|t} = f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t)$, and $e_{t+1|t} = \epsilon_{t+1}$, (P2)

- $\hat{y}_{t+2|t} = f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \cdots, y_{t-d+2}) = f_{\boldsymbol{\theta}_{t+2}}(\hat{y}_{t+1|t}, y_t, \cdots, y_{t+2-d})$,
- $\hat{y}_{t+h|t} = f_{\boldsymbol{\theta}_{t+h}}(\hat{y}_{t+h-1|t}, \cdots, \hat{y}_{t+1|t}, y_t, \cdots, y_{t+h-d})$ for $h > 2$, and, without a loss of generality, set $h < d$.

We can write

$$
\begin{aligned}
y_{t+h} =& f_{\boldsymbol{\theta}_{t+h}}(y_{t+h-1}, \cdots, y_{t+h-d}) + \epsilon_{t+h} \\
=& f_{\boldsymbol{\theta}_{t+h}}(\hat{y}_{t+h-1|t} + e_{t+h-1|t}, \cdots, \hat{y}_{t+2|t} + e_{t+2|t}, \hat{y}_{t+1|t} + e_{t+1|t}, y_t, \cdots, y_{t+h-d}) + \epsilon_{t+h} \\
\underset{te}{\approx}& f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a}) + \mathrm{D}\, f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a}) + \epsilon_{t+h} \\
=& f_{\boldsymbol{\theta}_{t+h}}(\hat{y}_{t+h-1|t}, \cdots, \hat{y}_{t+2|t}, \hat{y}_{t+1|t}, y_t, \cdots, y_{t+h-d}) \\
& + e_{t+h-1|t} \frac{\partial f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})}{\partial x_1} + \cdots + e_{t+2|t} \frac{\partial f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})}{\partial x_{h-2}} + e_{t+1|t} \frac{\partial f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})}{\partial x_{h-1}} \\
& + \epsilon_{t+h} \\
=& \hat{y}_{t+h|t} + e_{t+h|t},
\end{aligned}
$$

where $\boldsymbol{x} = (\hat{y}_{t+h-1|t} + e_{t+h-1|t}, \cdots, \hat{y}_{t+2|t} + e_{t+2|t}, \hat{y}_{t+1|t} + e_{t+1|t}, y_t, \cdots, y_{t+h-d})$, te means the first order Taylor series expansion of the function $f_{\boldsymbol{\theta}_{t+h}}$ at the point $\boldsymbol{a} = (\hat{y}_{t+h-1|t}, \cdots, \hat{y}_{t+2|t}, \hat{y}_{t+1|t}, y_t, \cdots, y_{t+h-d})$, $\mathrm{D}\, f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})$ denotes the matrix of partial derivatives, and $\frac{\partial}{\partial x_i}$ denotes the the partial derivative with respect to the ith argument in $f_{\boldsymbol{\theta}_{t+h}}$.

So we have

$$
\begin{aligned}
e_{t+h|t} &= e_{t+h-1|t} \frac{\partial f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})}{\partial x_1} + \cdots + e_{t+2|t} \frac{\partial f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})}{\partial x_{h-2}} + e_{t+1|t} \frac{\partial f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})}{\partial x_{h-1}} + \epsilon_{t+h}, \\
\text{and } e_{t+1|t} &= \epsilon_{t+1},
\end{aligned}
$$

which indicates that, **for optimal forecasts from a common forecast origin t, the h-step ahead forecast error, $e_{t+h|t}$ is functionally dependent on the past $h-1$ step ahead forecast errors,** $e_{t+1|t}, \cdots, e_{t+h-1|t}$.

2 References

Diebold, Francis X. 2017. *Forecasting.* Department of Economics, University of Pennsylvania. http://www.ssc.upenn.edu/~fdiebold/Textbooks.html.

Harvey, David, Stephen Leybourne, and Paul Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13 (2): 281–91.

Sommer, Benedikt. 2023. "Forecasting and Decision-Making for Empty Container Repositioning." PhD thesis, Technical University of Denmark.