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 (Diebold 2017)

- 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 (Sommer 2023)

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

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

  where te means the first order Taylor series expansion of the function $f_{\boldsymbol{\theta}_{t+2}}$ at the point $(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t), y_t, \ldots, y_{t-d+2})$. 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

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

  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, where the MA coefficients are complicated functions of observed data and unobserved model coefficients when f is non-linear.

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

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

where, we have $\boldsymbol{x}=\Big(\hat{y}_{t+h-1|t}+e_{t+h-1|t}, \cdots, \hat{y}_{t+2|t}+e_{t+2|t}, \hat{y}_{t+1|t}+e_{t+1|t}, y_t, \cdots, y_{t+h-d}\Big)$, $\boldsymbol{a}=(\hat{y}_{t+h-1|t}, \cdots, \hat{y}_{t+2|t}, \hat{y}_{t+1|t}, y_t, \cdots, y_{t+h-d})$, te means the first order Taylor series expansion of the function

$f_{\boldsymbol{\theta}_{t+h}}$ at the point $\boldsymbol{a}$, $\mathrm{D}\, f_{\boldsymbol{\theta}_{t+h}}(\boldsymbol{a})$ 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

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

which indicates that, **(P4) 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, $e_{t+1|t}, \cdots, e_{t+h-1|t}$.**

1.4 Extension to include exogenous variables

Then we try to extend the dependence structure to include exogenous variables. Assuming that a time series $y_1, \cdots, y_T$ is generated by the non-stationary autoregressive process with exogenous variables

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

where $\boldsymbol{u}_t = (u_{1,t}, \cdots, u_{k,t})'$.

Also write

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

where

- $\hat{y}_{t+1|t} = f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t, \boldsymbol{u}_t)$, and $e_{t+1|t} = \epsilon_{t+1}$, (P2)
- $\hat{y}_{t+2|t} = f_{\boldsymbol{\theta}_{t+2}}(f_{\boldsymbol{\theta}_{t+1}}(\boldsymbol{x}_t, \boldsymbol{u}_t), y_t, \cdots, y_{t-d+2}) = f_{\boldsymbol{\theta}_{t+2}}(\hat{y}_{t+1|t}, y_t, \cdots, y_{t+2-d})$,
- $\hat{y}_{t+h|t} = f_{\boldsymbol{\theta}_{t+h}}(\hat{y}_{t+h-1|t}, \cdots, \hat{y}_{t+1|t}, y_t, \cdots, y_{t+h-d})$ for $h > 2$, and, without a loss of generality, set $h < d$.

We can write

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

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

Thus, P3 and P4 still hold when exogenous variables are included in the autoregressive framework. However, the MA coefficients for the MA(h-1) process (for P3) and the regression coefficients (for P4) now also depend on $\boldsymbol{u}_t, \cdots, \boldsymbol{u}_{t+h-1}$.

2 PID generalization

The conformal PID controller is given by

$$q_{t+1} = \underbrace{\eta(\mathrm{err}_t - \alpha)}_{\mathrm{P}} + \underbrace{r_t\left(\sum_{i=1}^{t}(\mathrm{err}_i - \alpha)\right)}_{\mathrm{I}} + \underbrace{g'_t}_{\mathrm{D}}.$$

The idea to generalize the method is that, for each forecast horizon h, the iteration is given by

$$q_{t+h|t} = \underbrace{\eta_h(\mathrm{err}_{t|t-h} - \alpha)}_{\mathrm{P}} + \underbrace{r_t\left(\sum_{i=h+1}^{t} w_i(\mathrm{err}_{i|i-h} - \alpha)\right)}_{\mathrm{I}} + \underbrace{\hat{q}_{t+h|t}}_{\mathrm{D}}, \text{ for } t > h,$$

Let $q_{t+h|t} = e_{t+h|t}$, $q_{t+h|t}$ follows a MA$(h-1)$ process, and it is functionally dependent on the past $h-1$ step ahead forecast errors, i.e., $e_{t+h-1|t}, \cdots, e_{t+1|t}$.

So, $\hat{e}_{t+h|t}$ can be a forecast combination of the MA$(h-1)$ model fitted using $e_{1+h|1}, \cdots, e_{t-1+h|t-1}$ and **a linear regression (seldom reliable extrapolation)** of $e_{t+h|t}$ on $e_{t+1|t}, \cdots, e_{t+h-1|t}$. Perhaps dynamic regression model can be used to replace linear regression. For $h = 1$, the snaïve method can be used to produce forecasts.

The above statement means that $e_{t+h|t}$ is stationary. However, in practice, it is hard to achieve stationary forecast errors especially for time series with trend and seasonality.

Checklist

- ☐ P4 and its derivation
- ☐ design of the scorecaster $\hat{q}_{t+h|t}$, MA$(h-1)$+LR
- ☐ stationary $e_{t+h|t}$

3 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.