# Time Series Analysis 2
# DS351

# Exponential smoothing

# Motivation

We can forecast simply using the previous value:

$$\hat{y}_{T+1} = y_T$$

# Motivation

We can forecast simply using the previous value:

$$\hat{y}_{T+1} = y_T$$

or a simple average

$$\hat{y}_{T+1} = \frac{1}{T} \sum_{i=1}^{T} y_t$$

# Motivation

We can forecast simply using the previous value:

$$\hat{y}_{T+1} = y_T$$

or a simple average

$$\hat{y}_{T+1} = \frac{1}{T} \sum_{i=1}^{T} y_t$$

- Notice that both forecasts use **weighted average** of previous observations.

# Motivation

We can forecast simply using the previous value:

$$\hat{y}_{T+1} = y_T$$

or a simple average

$$\hat{y}_{T+1} = \frac{1}{T}\sum_{i=1}^{T} y_t$$

- Notice that both forecasts use **weighted average** of previous observations.
- We want to make a forecasting model that lie between these two extremes.

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \\ + \alpha^{T-1} y_1 + \alpha^T l_0$$

This model has two data-dependent parameters:

- $\alpha$ is the **smoothing parameter**
- $l_0$ is the **initial value**

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \\ + \alpha^{T-1} y_1 + \alpha^T l_0$$

Observations:

- $\alpha < \alpha(1-\alpha) < \alpha(1-\alpha)^2 < \dots$ (decreasing weights)

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots$$
$$+ \alpha^{T-1} y_1 + \alpha^T l_0$$

Observations:

- $\alpha < \alpha(1-\alpha) < \alpha(1-\alpha)^2 < \dots$ (decreasing weights)
- The sum of the weights is:

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \\ + \alpha^{T-1} y_1 + \alpha^T l_0$$

The forecast at $T = 2$ is

$$\hat{y}_2 = \alpha y_1 + \alpha(1-\alpha) l_0$$

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \\ + \alpha^{T-1} y_1 + \alpha^T l_0$$

The forecast at $T = 2$ is

$$\hat{y}_2 = \alpha y_1 + \alpha(1-\alpha) l_0$$

The forecast at $T = 3$ is

$$\hat{y}_3 = \alpha y_2 + \alpha(1-\alpha) y_1 + \alpha(1-\alpha)^2 l_0$$

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \\ + \alpha^{T-1} y_1 + \alpha^T l_0$$

The forecast at $T = 2$ is

$$\hat{y}_2 = \alpha y_1 + \alpha(1-\alpha) l_0$$

The forecast at $T = 3$ is

$$\begin{aligned} \hat{y}_3 &= \alpha y_2 + \alpha(1-\alpha) y_1 + \alpha(1-\alpha)^2 l_0 \\ &= \alpha y_2 + (1-\alpha)\hat{y}_1 \end{aligned}$$

# Exponential smoothing

Idea: give the largest weight to the most recent:

$$\hat{y}_{T+1} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \dots \\ + \alpha^{T-1} y_1 + \alpha^T l_0$$

The forecast at $T = 2$ is

$$\hat{y}_2 = \alpha y_1 + \alpha(1-\alpha) l_0$$

The forecast at $T = 3$ is

$$\hat{y}_3 = \alpha y_2 + \alpha(1-\alpha) y_1 + \alpha(1-\alpha)^2 l_0 \\ = \alpha y_2 + (1-\alpha)\hat{y}_1$$

at $T = 4$ is

$$\hat{y}_4 = \alpha y_3 + (1-\alpha)\hat{y}_2$$

# Exponential smoothing

Two forms of ES:

$$\hat{y}_{T+1} = \alpha y_T + (1-\alpha)\hat{y}_T$$
$$= \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0$$

where $\alpha, l_0$ is an **initial value**, are two parameters to be learned from the data $y_1, y_2, \ldots, y_T$

# Learning parameters from the data

ES model:

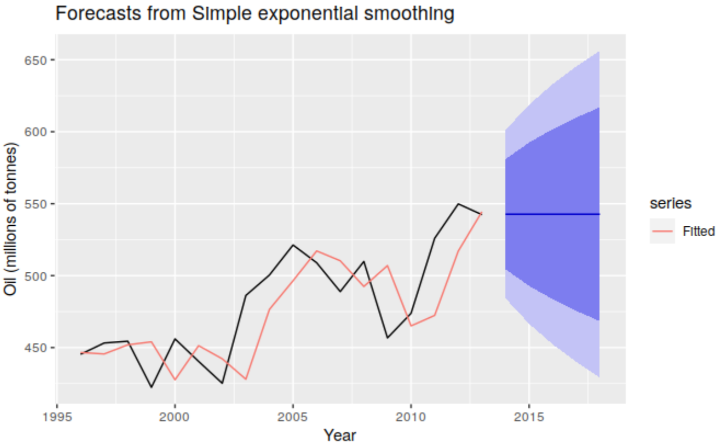
$$\hat{y}_{T+1} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0$$

From input data $y_1, y_2, \ldots, y_T$, we need to find $\alpha$ and $l_0$ that minimize the SSE.

$$\text{SSE}(\alpha, l_0) = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2$$

This is a function of $\alpha, \alpha^2, \ldots, \alpha^T, l_0$, so not as easy to optimize as linear regression

# Example: oil production in Saudi Arabia




Learned ES parameters: $\hat{\alpha} = 0.83$ and $\hat{l}_0 = 446.6$.

# Holt's linear trend method

When there is a trend but no seasonality, use Holt's method **Holt's method**

$$
\begin{aligned}
\text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\
\text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1},
\end{aligned}
$$

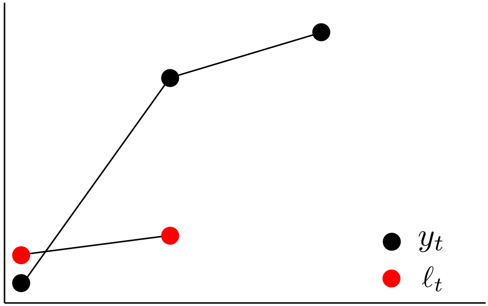
There are **4** parameters here: $\alpha$, $\beta$, $\ell_0$ and $b_0$.

# Holt's linear trend method

**Holt's method**

$$
\begin{aligned}
\text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + h b_t \\
\text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1},
\end{aligned}
$$

- $l_t$ is the level (estimate of $y_t$).
- $b_t$ is the slope.
- Suppose that we have $l_{t-1}$ and $b_{t-1}$.

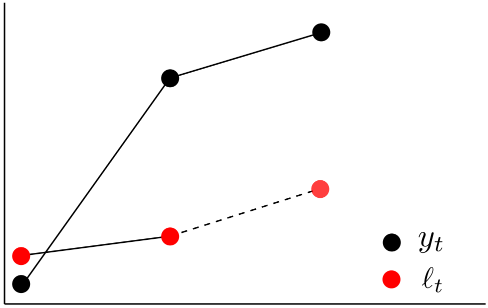


# Holt's linear trend method

**Holt's method**

$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\ \text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ \text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \end{aligned}$$

- $l_t$ is the "average" between $y_t$ and $l_{t-1} + b_{t-1}$.
- Find $l_{t-1} + b_{t-1}$.

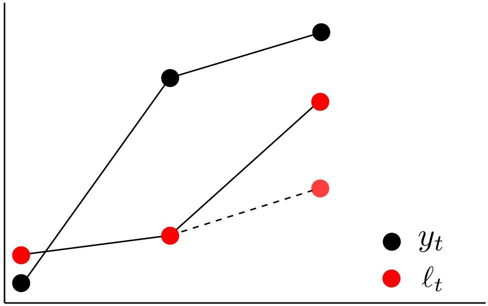


# Holt's linear trend method

**Holt's method**

$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\ \text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ \text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \end{aligned}$$

- $l_t$ is the "average" between $y_t$ and $l_{t-1} + b_{t-1}$.
- Find $l_{t-1} + b_{t-1}$.
- Then find $l_t$.

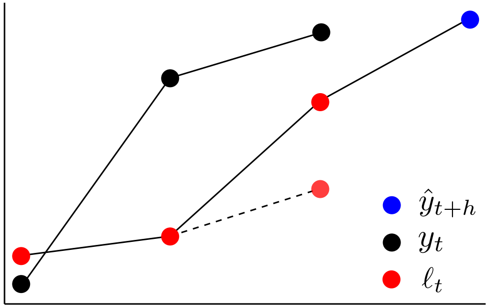


# Holt's linear trend method

**Holt's method**

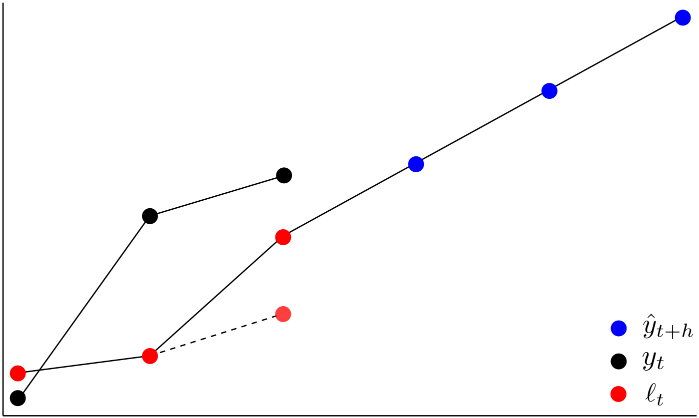
$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\ \text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ \text{Trend equation} \quad & b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}, \end{aligned}$$

- $b_t$ is the "average" between $l_t - l_{t-1}$ and $b_{t-1}$.
- Start the first forecast $\hat{y}_{t+1|t}$.



# Holt's linear trend method

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$




The forecast is a linear function of $h$.

# Air passengers data

| Year | Time | Observation | Level | Slope | Forecast |
|---|---|---|---|---|---|
| | $t$ | $y_t$ | $\ell_t$ | $b_t$ | $y_{t\|t-1}$ |
| 1989 | 0 | | 15.57 | 2.102 | |
| 1990 | 1 | 17.55 | 17.57 | 2.102 | 17.67 |
| 1991 | 2 | 21.86 | 21.49 | 2.102 | 19.68 |
| 1992 | 3 | 23.89 | 23.84 | 2.102 | 23.59 |
| 1993 | 4 | 26.93 | 26.76 | 2.102 | 25.94 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 2016 | 27 | 72.60 | 72.50 | 2.102 | 72.02 |
| | $h$ | | | | $\hat{y}_{t+h\|t}$ |
| | 1 | | | | 74.60 |
| | 2 | | | | 76.70 |
| | 3 | | | | 78.80 |
| | 4 | | | | 80.91 |
| | 5 | | | | 83.01 |

# Damped Holt's method

- Linear trend is not realistic in many situations.
- Examples: Total factory output with a fixed number of machine.

**Damped Holt's method** (Gardner & McKenzie, 1985)

Fix $0 \le \phi \le 1$

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t \\
\ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}.
\end{aligned}$$

# Damped Holt's method

- Linear trend is not realistic in many situations.
- Examples: Total factory output with a fixed number of machine.

**Damped Holt's method** (Gardner & McKenzie, 1985)

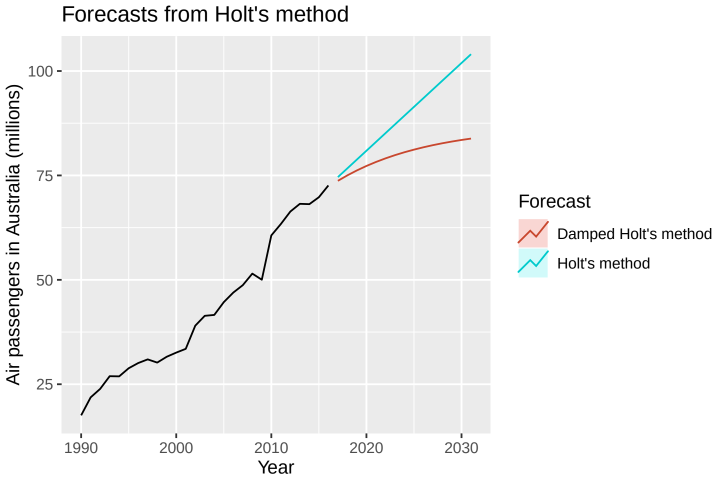
Fix $0 \leq \phi \leq 1$

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + (\phi + \phi^2 + \cdots + \phi^h) b_t \\ \ell_t &= \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}. \end{aligned}$$

- $\phi = 1 \rightarrow$ Holt's method
- $\phi = 0 \rightarrow$ forecast with a constant
- In practice, $\phi \geq 0.8$.

# Air passengers data

$$\phi = 0.9$$



# Holt-Winters' seasonal method

- Use this method when there is seasonality.

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{\text{t+h-m}}$$
$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},$$

- Basically Holt's method + seasonality.
- $m$ is the frequency of seasonality e.g. $m = 12$ for monthly data.

# Holt-Winters' seasonal method

- Use this method when there is seasonality.

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{\text{t+h-m}} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}
$$

- Basically Holt's method + seasonality.
- $m$ is the frequency of seasonality e.g. $m = 12$ for monthly data.
- $l_t$ is the "average" between **observation with seasonality removed** $y_t - s_{t-m}$ and $l_{t-1} + b_{t-1}$.

# Holt-Winters' seasonal method

- Use this method when there is seasonality.

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{\text{t+h-m}}$$
$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},$$

- Basically Holt's method + seasonality.
- $m$ is the frequency of seasonality e.g. $m = 12$ for monthly data.
- $l_t$ is the "average" between **observation with seasonality removed** $y_t - s_{t-m}$ and $l_{t-1} + b_{t-1}$.
- $s_t$ is the "average" between **observation with level and trend removed** $y_t - l_{t-1} - b_{t-1}$ and the value of previous season $s_{t-m}$.

# Holt-Winters' seasonal method

**Holt-Winters' method**

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}$$

- $s_{t+h-m}$ the latest seasonality in the data that has the same seasonal index (month, day of the week etc.) as $t+h$.

# Holt-Winters' seasonal method

**Holt-Winters' method**

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}
$$

- $s_{t+h-m}$ the latest seasonality in the data that has the same seasonal index (month, day of the week etc.) as $t+h$.
- For example, if $t =$ January, 2019, $h = 2$, and $m = 12$, then $t + h =$ March, 2019 and $t + h - m =$ March, 2018.

# Holt-Winters' seasonal method

**Holt-Winters' method**

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}$$

- $s_{t+h-m}$ the latest seasonality in the data that has the same seasonal index (month, day of the week etc.) as $t + h$.
- For example, if $t =$ January, 2019, $h = 2$, and $m = 12$, then $t + h =$ March, 2019 and $t + h - m =$ March, 2018.
- There are a lot of parameters now: $\alpha$, $\beta$, $\gamma$, $\ell_0$, $b_0$, $s_{-m+1}$, $s_{-m+2}, \ldots, s_0$.

# Holt-Winters' multiplicative method

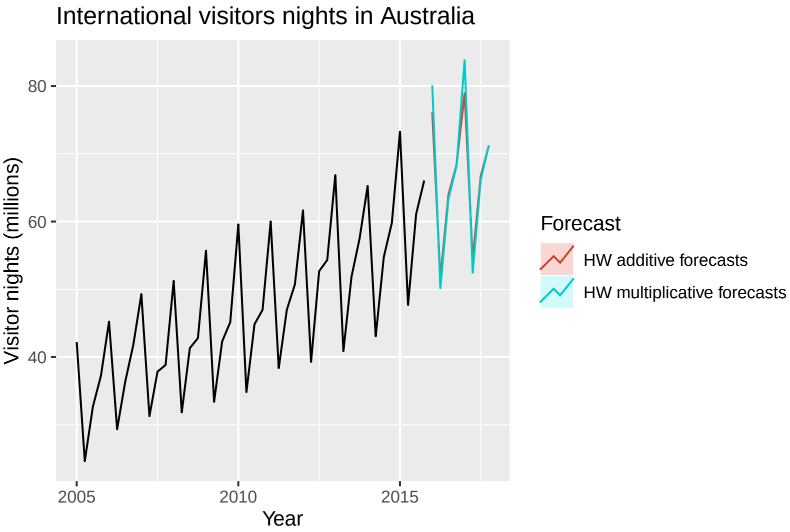
We can replace

$$\text{add by } s_t \rightarrow \text{multiply by} s_t$$
$$\text{subtract by } s_t \rightarrow \text{divide by} s_t.$$

**Holt-Winters' multiplicative method**

$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t-} \\
\ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}.
\end{aligned}$$

# International visitors nights in Australia



# Forecasts using Holt-Winters' method

| | $t$ | $y_t$ | $\ell_t$ | $b_t$ | $s_t$ | $y_t$ |
|---|---|---|---|---|---|---|
| 2004 Q1 | -3 | | | | 9.70 | |
| 2004 Q2 | -2 | | | | -9.31 | |
| 2004 Q3 | -1 | | | | -1.69 | |
| 2004 Q4 | 0 | | 32.26 | 0.70 | 1.31 | |
| 2005 Q1 | 1 | 42.21 | 32.82 | 0.70 | 9.50 | 42.66 |
| 2005 Q2 | 2 | 24.65 | 33.66 | 0.70 | -9.13 | 24.21 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 2015 Q4 | 44 | 66.06 | 63.22 | 0.70 | 2.35 | 64.22 |
| | $h$ | | | | | $y_{t+h\|t}$ |
| 2016 Q1 | 1 | | | | | 76.10 |
| 2016 Q2 | 2 | | | | | 51.60 |
| 2016 Q3 | 3 | | | | | 63.97 |
| 2016 Q4 | 4 | | | | | 68.37 |
| 2017 Q1 | 5 | | | | | 78.90 |
| 2017 Q2 | 6 | | | | | 54.41 |