Stat 443. Time Series and Forecasting.

Key ideas: additive stochastic models associated with forecasting rules;
forecast is the conditional expectation of future observation given the observed past;
Holt linear exponential smoothing.

Two recursion equations: one for level and one for slope b (called trend in R documentation, but is different from trend in STL).

The recursions and h-step forecasts are as follows.

$$\begin{aligned}
\widehat{\ell}_t &= \alpha y_t + (1-\alpha)(\widehat{\ell}_{t-1} + \widehat{b}_{t-1}) = \widehat{\ell}_{t-1} + \widehat{b}_{t-1} + \alpha(y_t - \widehat{\ell}_{t-1} - \widehat{b}_{t-1}) \\
\widehat{b}_t &= \beta(\widehat{\ell}_t - \widehat{\ell}_{t-1}) + (1-\beta)\widehat{b}_{t-1} = \widehat{b}_{t-1} + \beta(\widehat{\ell}_t - \widehat{\ell}_{t-1} - \widehat{b}_{t-1}) \\
\widehat{y}_{t+1|t} &= \widehat{\ell}_t + \widehat{b}_t \\
\widehat{y}_{t+h|t} &= \widehat{\ell}_t + h\widehat{b}_t, \quad h = 1, 2, \ldots
\end{aligned}$$

The smoothed level $\widehat{\ell}_t$ is a convex combination of the *most recent observation* and the *local linear projection of the previous smoothed value* .
The smoothed slope $\widehat{b}_t$ is a convex combination of the *most recent slope change* and the *previous smoothed slope*.
The h-step forecast is based on a constant slope using the last smoothed slope value. (Similar to linear extrapolation).

α, β are estimated by minimizing the in-sample root mean square prediction error.

If the HW implementation, such as **`forecast::holt`**, outputs the 1-step to 10-step forecasts $\widehat{y}_{t+h|t}$ at the end of the series, but not $\widehat{\ell}_t$ and $\widehat{b}_t$, then

$$\begin{aligned}
\widehat{y}_{t+1|t} &= \widehat{\ell}_t + \widehat{b}_t \\
\widehat{y}_{t+2|t} &= \widehat{\ell}_t + 2\widehat{b}_t \\
\widehat{b}_t &= \widehat{y}_{t+2|t} - \widehat{y}_{t+1|t} \\
\widehat{\ell}_t &= \widehat{y}_{t+1|t} - \widehat{b}_t
\end{aligned}$$

$\widehat{\ell}_t$ and $\widehat{b}_t$ are needed for the moving 1-step ahead forecasts for getting holdout set forecast errors.

Pseudo-code for out-of-sample rmse (linear exponential smoothing)
Part 1:

- Input `train` with size n.
- Estimate α, β parameters as $\widehat{\alpha}$, $\widehat{\beta}$. and get the two smoothed series $\{\widehat{\ell}_t\}$ and $\{\widehat{b}_t\}$. $\widehat{\ell}_n$ is the last smoothed level value of the training set. $\widehat{b}_n$ is the last smoothed slope value of the training set (these are in the $\$$`coefficents` component of `HoltWinters` output).
- Save $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\ell}_n$, $\widehat{b}_n$.

Part 2: Separate out-of-sample rmse from linear exponential smoothing because R (HoltWinters and forecast::holt) and SAS might have different values of $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\ell}_n$, $\widehat{b}_n$.

- Input $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\ell}_n$, $\widehat{b}_n$, `holdout` with size $n_{holdout}$
- sse$\leftarrow$ 0
- fc$\leftarrow$ $\widehat{\ell}_n + \widehat{b}_n$; yt$\leftarrow$ holdout[1]; newfc$\leftarrow$ fc; fcvec[1]$\leftarrow$ fc; fcerror$\leftarrow$ yt-fc; sse$\leftarrow$ sse$+$ fcerror^2; $\ell_{prev} \leftarrow \widehat{\ell}_n$; $b_{prev} \leftarrow \widehat{b}_n$.
- for i in $2, \ldots, n_{holdout}$:
- $\ell_{new} \leftarrow \widehat{\alpha}\times$ holdout[i-1] $+(1-\widehat{\alpha})\times$ fc; $b_{new} \leftarrow \widehat{\beta}(\ell_{new} - \ell_{prev}) + (1-\widehat{\beta})b_{prev}$; fc$\leftarrow \ell_{new} + b_{new}$; fcvec[i]$\leftarrow$ fc; yt$\leftarrow$ holdout[i]; fcerror$\leftarrow$ yt-fc; sse$\leftarrow$ sse$+$ fcerror^2;
- $\ell_{prev} \leftarrow \ell_{new}$; $b_{prev} \leftarrow b_{new}$.
- end for
- return rmse=sqrt(sse/$n_{holdout}$)

On the previous page, why do R HoltWinters, forecast::holt and SAS have (slightly) different values of $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\ell}_n$, $\widehat{b}_n$.

Qualitatively, forecasts are not affected by the implementation of exponential smoothing rules.

Stochastic model for linear exponential smoothing (upper case Y, L, B and ϵ for random variables). Write a stochastic model as:

$$\begin{aligned}
L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + B_{t-1}) = L_{t-1} + B_{t-1} + \alpha(Y_t - L_{t-1} - B_{t-1}) \\
B_t &= \beta(L_t - L_{t-1}) + (1-\beta)B_{t-1} = B_{t-1} + \beta(L_t - L_{t-1} - B_{t-1}) \\
Y_{t+1} &= L_t + B_t + \epsilon_{t+1}, \quad \epsilon_{t+1} \text{ random innovation with mean 0}
\end{aligned}$$

Then the idea is to use differencing to eliminate $\{L_t\}$ and $\{B_t\}$, and represent $\{Y_t\}$ in terms of $\{\epsilon_t\}$.

$$\begin{aligned}
Y_t &= L_{t-1} + B_{t-1} + \epsilon_t \text{ or } Y_t - L_{t-1} - B_{t-1} = \epsilon_t \\
Y_{t+1} - Y_t &= [L_t - L_{t-1}] + [B_t - B_{t-1}] + [\epsilon_{t+1} - \epsilon_t] \\
&= [B_{t-1} + \alpha(Y_t - L_{t-1} - B_{t-1})] + \beta(L_t - L_{t-1} - B_{t-1}) + \epsilon_{t+1} - \epsilon_t \\
&= [B_{t-1} + \alpha\epsilon_t] + \beta\alpha\epsilon_t + \epsilon_{t+1} - \epsilon_t \\
&= B_{t-1} + \epsilon_{t+1} - (1 - \alpha - \alpha\beta)\epsilon_t \\
Y_t - Y_{t-1} &= B_{t-2} + \epsilon_t - (1 - \alpha - \alpha\beta)\epsilon_{t-1} \\
\Delta_2 Y_{t+1} &:= (Y_{t+1} - Y_t) - (Y_t - Y_{t-1}) \\
&= [B_{t-1} - B_{t-2}] + [\epsilon_{t+1} - (1 - \alpha - \alpha\beta)\epsilon_t] - [\epsilon_t - (1 - \alpha - \alpha\beta)\epsilon_{t-1}] \\
&= \beta\alpha\epsilon_{t-1} + \epsilon_{t+1} - (2 - \alpha - \alpha\beta)\epsilon_t + (1 - \alpha - \alpha\beta)\epsilon_{t-1} \\
&= \epsilon_{t+1} - (2 - \alpha - \alpha\beta)\epsilon_t + (1 - \alpha)\epsilon_{t-1}
\end{aligned}$$

Pay attention to the technique. It can be use to convert other exponential smoothing rules into stochastic models.

Second difference is a linear function of three consecutive ϵ's. This is an example of an ARIMA model, whose theory leads to forecast standard errors. Initially Holt and Winters proposed simle intuitive forecasting rules, and later others deduced appropriate forecast standard errors.

Homework (webwork+upload derivation). Damped trend (slope) exponential smoothing
Convert to simple representation similar to preceding.

The damping parameter is denoted as ϕ; $\phi \in [0,1]$. This allows for sublinear extrapolaton for forecasts. Recursion equations are:

$$
\begin{aligned}
\widehat{\ell}_t &= \alpha y_t + (1-\alpha)(\widehat{\ell}_{t-1} + \phi \widehat{b}_{t-1}) \\
\widehat{b}_t &= \beta(\widehat{\ell}_t - \widehat{\ell}_{t-1}) + (1-\beta)\phi \widehat{b}_{t-1} \\
\widehat{y}_{t+1|t} &= \widehat{\ell}_t + \phi \widehat{b}_t \\
\widehat{y}_{t+h|t} &= \widehat{\ell}_t + \widehat{b}_t \textstyle\sum_{i=1}^{h} \phi^i = \widehat{\ell}_t + \widehat{b}_t C_h, \quad h = 1, 2, \ldots, \\
& C_h = \textstyle\sum_{i=1}^{h} \phi^i = (1-\phi^{h+1})/(1-\phi) - 1.
\end{aligned}
$$