Time Series Lecture 4: Exponential Smoothing

CMU MSP 36618 March 3, 2016

Hyndman sections 7.1-7.5

I. **(Additive) Simple (a.k.a. single) Exponential Smoothing (SES)**

a. "Naïve" forecasts center on the latest available data point, putting all of the "weight" on the last value in the time series.

b. Forecasts for the "average" method center on the mean of all data, spreading the weight equally across all T time points.

c. Exponential smoothing is based on the reasonable idea that forecasts are best made using ***decreasing weights*** some distance back in time. Because the weights are exponentially decreasing with a parameter α, they eventually become negligible if we go back far enough in time, with the rate depending on α.

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

Note that this uses a standard algebraic sequence $\sum_{t=1}^{\infty} \alpha(1-\alpha)^{t-1}$ that adds to 1, so it is (nearly) a weighted average unless α is small (slow decay) and/or T is small (insufficient terms).

d. E.g., the sum of the weights reaches 0.9 over the last 11 values for α=0.2, over 5 values for α=0.4, 3 values for α=0.6, and 2 values for α=0.8.

e. Equivalent representations of SES

i. Weighted average form

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Interpretation: The next forecast value, at time t+1, when we have collected data up to time t, is a weighted average of the actual value at time t, y_t, and the forecast for time t, $\hat{y}_{t|t-1}$, that we made "last time" (one time unit back), with weights α on the observed value and (1-α) on the forecast value.

To get the whole thing started, i.e., to forecast at time t=2 after we have observed at time t=1, we use $\hat{y}_{2|1} = \alpha y_1 + (1-\alpha)\hat{y}_{1|0} = \alpha y_1 + (1-\alpha) l_o$ where we define l_0 to be some starting forecast before observing any data.

Iteratively: $\hat{y}_{3|2} = \alpha y_2 + (1-\alpha)\hat{y}_{2|1} = \alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2 l_o$

If we keep going we get the equation of Ic above (but with t starting at 0 instead of 1 and with a final term of $\alpha(1-\alpha)^T l_0$). The farther we go forward in time, the less the choices of l_o matters.

ii. Component form

Forecast Equation: $\hat{y}_{t+1|t} = l_t$

Smoothing Equation: $l_t = \alpha\, y_t + (1-\alpha) l_{t-1}$

Interpretation: Signal + noise decomposition gives $y_t = l_t + e_t$. With smoothing, the estimate of the signal at time t is a weighted average of the current data value and the signal at time t-1. We assume that the signal is changing, but not in any consistent, predictable way. So we forecast that the next value is equal to the current signal estimate.

In models with a trend, the Forecast Equation is no longer trivial.

iii. Error correction form

Forecast Equation: $\hat{y}_{t+1|t} = l_t$

Smoothing Equation: $l_t = l_{t-1} + \alpha(\, y_t \; - \; l_{t-1})$

$$= \; l_{t-1} + \alpha\, e_t \qquad e_t = \; (\, y_t \; - \; l_{t-1}) = \; y_t - \; \hat{y}_{t|t-1}$$

iv. Estimation: Maximum likelihood estimation of α and l_0 is the default in the forecast package, but other choices such as minimal MSE are possible.

v. **Forecasting**: Means of forecasts come directly from these equations. These can be iterated to give forecasts "h" steps into the future by treating the h-1 earlier forecast values as known values. Forecast intervals take into account the full uncertainty (using "state-space models" with the error correction form equations).

vi. **R package "forecast"**:

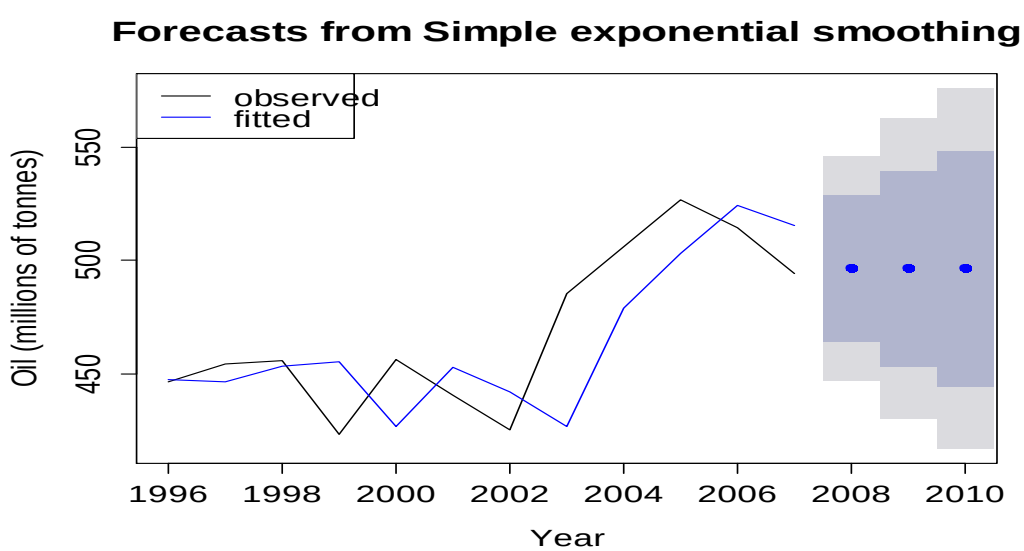
ets() fits models with various forms for Error, Trend, and Seasonality.

forecast() takes a model fit and makes forecasts (or it can run ets() first if data is given instead of an ets model object.)

`ses(x, h=10)` is a "wrapper" function that runs `forecast(, h=10)` on `ets(x, model="ANN")` where x is the time series objects, and h is the "horizon" for prediction. "ANN" is additive error, no trend, no seasonality.

Example: Oil production in Saudi Arabia

```
library(fpp) # for data sets
mSes <- ses(oildata, h=3)
summary(mSes)
# Forecast method: Simple exponential smoothing
#
# Model Information:
# ETS(A,N,N)
#
#   Smoothing parameters: alpha = 0.8921
#
#   Initial states:    l = 447.4808
#   sigma:  25.1221
#      AIC     AICc      BIC
# 111.1888 112.5221 112.1586
#
# Error measures:
#                    ME     RMSE      MAE       MPE ...
# Training set 4.578526 25.12207 20.05797 0.8091703 ...
#
# Forecasts:
#      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
# 2008       496.4923 464.2971 528.6876 447.2540 545.7307
# 2009       496.4923 453.3488 539.6359 430.5100 562.4747
# 2010       496.4923 444.6638 548.3208 417.2275 575.7572

plot(mSes, ylab="Oil (millions of tonnes)", xlab="Year")
lines(fitted(mSes), col="blue", type="l")
legend("topleft", c("observed","fitted"), lty=1, col=c(1,4))
```



vii. **SAS ESM (Exponential Smoothing Models)**:

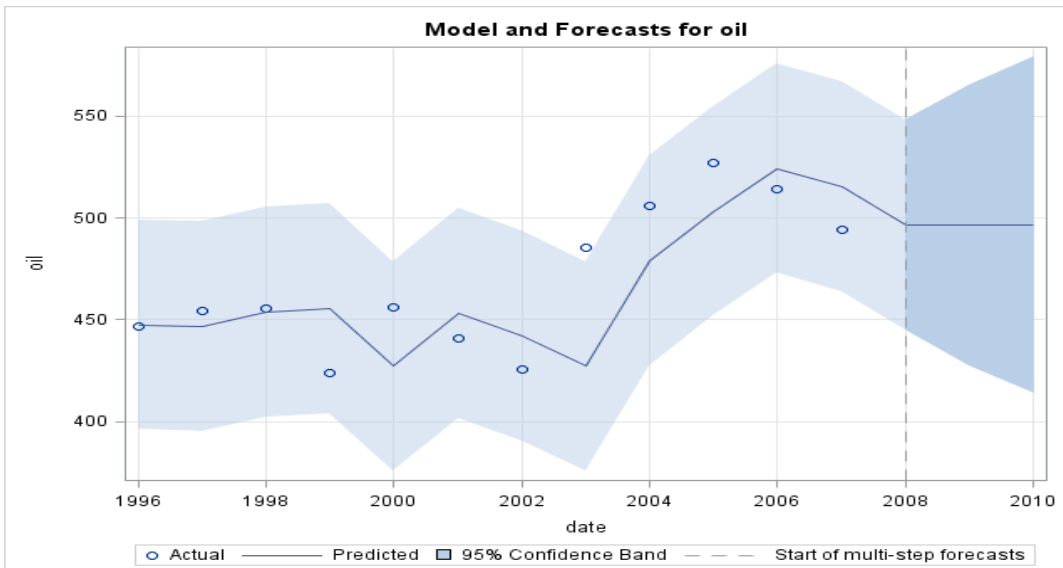
```
proc esm data=<input-data-set> out=<output-data-set>
         back=<n> lead=<n> ;
    id <time-ID-variable> interval=<frequency>;
    forecast <time-series-variables>;
run;
```

Example: Oil production in Saudi Arabia

```
PROC ESM DATA=oil OUT=oilPred PRINT=(ESTIMATES FORECASTS)
     LEAD=3 PLOT=(MODELFORECASTPLOT);
  ID date interval = YEAR;
  FORECAST oil / MODEL=SIMPLE;
RUN;
```

Simple Exponential Smoothing Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Level Weight	0.89211	0.21822	4.09	0.0018



II. Holt's Exponential Smoothing with a linear trend (double exponential smoothing)

a. Component Equations predicting forward horizon "h" time units (with additive errors)

Forecast Equation: $\hat{y}_{t+h|t} = l_t + hb_t$

Level Equation: $l_t = \alpha\, y_t + (1-\alpha)(l_{t-1} + b_{t-1})$

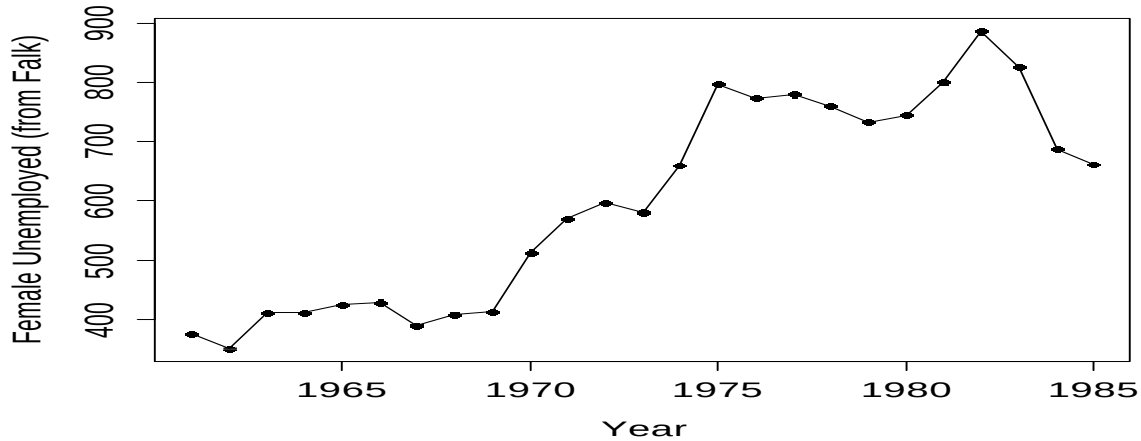
Trend Equation: $b_t = \beta^*\,(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$

(Notes: $\Delta y = \Delta x$ * slope → Δy = slope. Programs may show $\beta = \alpha\beta^*$.)

b. Holt's method in R

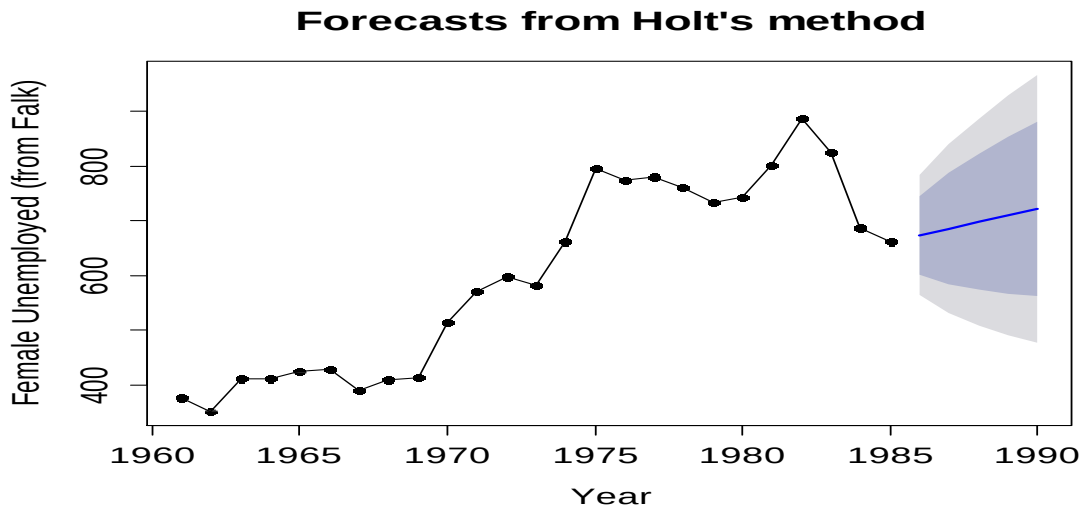
```
# Holt's method for female unemployment
fun = scan("data/female.txt", quiet=TRUE) # From Falk
# Convert monthly data to yearly data (non-seasonal data)
fun = ts(apply(matrix(fun, nrow=12), MARGIN=2, FUN=mean),
         start=1961)
```

```
plot(fun, xlab="Year", ylab="Female Unemployed (from Falk)",
     type="o", pch=16)
```




```
library(forecast) # for holt(), ets(), etc.
mHolt = holt(fun, h=5)
print(summary(mHolt))
# Forecast method: Holt's method
# Model Information: ETS(A,A,N)
#
#   Smoothing parameters:
#     alpha = 0.9999
#     beta  = 1e-04
#
#   Initial states:
#     l = 359.745
#     b = 12.0023
#
#   sigma:  55.8512
#
#      AIC     AICc      BIC
# 289.6064 291.6064 294.4819
#
# Error measures:
#                      ME     RMSE      MAE     MAPE  ...
# Training set 0.06307834 55.85119 41.47882 6.756772  ...
#
# Forecasts:
#      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
# 1986       673.5895 602.0133 745.1657 564.1232 783.0558
# 1987       685.5919 584.3629 786.8210 530.7754 840.4084
# 1988       697.5943 573.6084 821.5803 507.9741 887.2146
# 1989       709.5968 566.4229 852.7706 490.6313 928.5623
# 1990       721.5992 561.5180 881.6804 476.7761 966.4223
```

```
plot(mHolt, xlab="Year", ylab="Female Unemployed (from Falk)",
     type="o", pch=16)
```



c. AIC is not directly obtainable from a "forecast" object, but it is from an "ets" object.

```
class(mHolt)  # [1] "forecast"
class(mHolt$model) # [1] "ets"
AIC(mHolt)
# Error in UseMethod("logLik") :
#   no applicable method for 'logLik' applied to an object of
#   class "forecast"
AIC(mHolt$model) # [1] 289.6064
AIC(ets(fun, model="ANN")) # [1] 286.6565
```

R note: defining

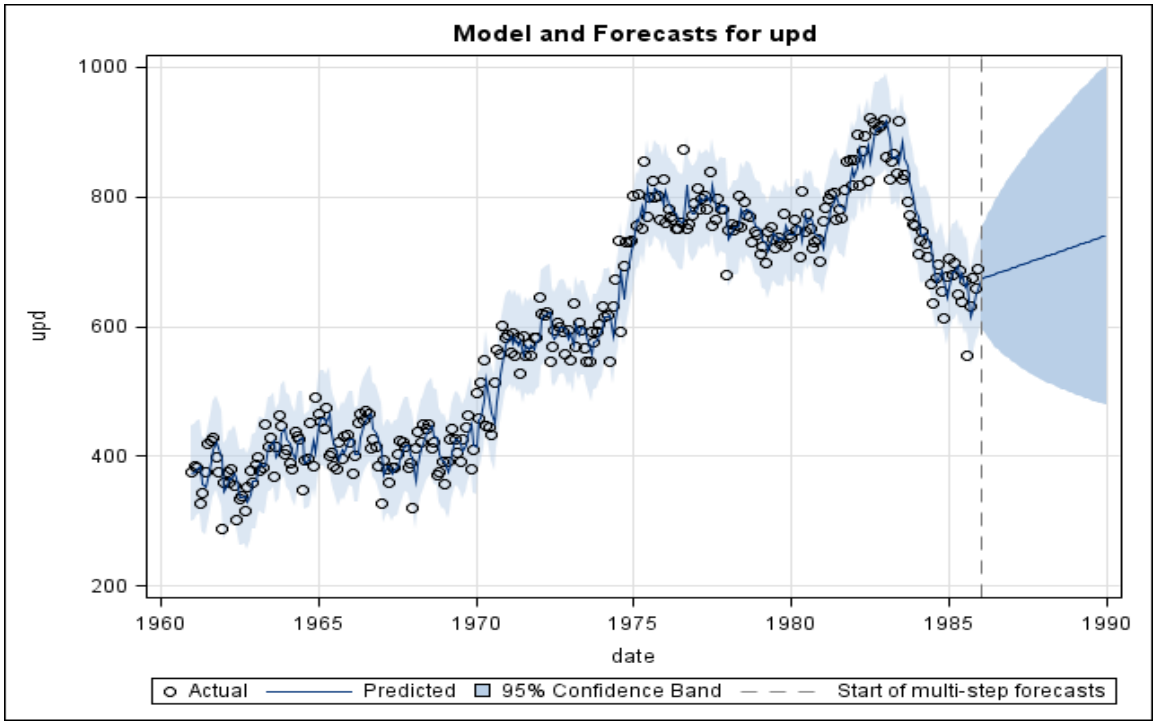
```
 AIC.forecast = function(object, ..., k=2) AIC(object$model)
```

allows AIC(mHolt) to work.

d. Holt's method in SAS (PROC ESM with "MODEL=LINEAR")

```
TITLE1 "Holt's Exponential Smoothing with a Trend";
TITLE2 'Unemployed Females Data';
DATA upd;
  INFILE female;
  INPUT upd @@;
  date = INTNX('month', '01jan61'd, _N_-1);
  FORMAT date yymon.;
RUN;

TITLE2 "Holt's Exponential Smoothing with a Trend";
PROC ESM DATA=upd OUT=updPred PRINT=(ESTIMATES FORECASTS)
     LEAD=48 PLOT=(MODELFORECASTS);
  ID date interval = MONTH;
  FORECAST upd / MODEL=LINEAR ALPHA=0.05;
  ODS OUTPUT FORECASTS=fcUpd;
RUN;
```

```
Linear Exponential Smoothing Parameter Estimates
                                  Standard                  Approx
Parameter          Estimate          Error     t Value     Pr > |t|
Level Weight        0.48564        0.03546       13.70       <.0001
Trend Weight      0.0010000      0.0053798        0.19       0.8527
```



III. Exponential smoothing with "dampening" of the trend

Taylor (International Journal of Forecasting, 2003, Vol. 19, pp. 715-725):
"The robustness and accuracy of exponential smoothing forecasting has led to its widespread use in applications where a large number of series necessitates an automated procedure, such as inventory control. Although Holt's method has tended to be the most popular approach for trending series, its linear forecast function has been criticised for tending to overshoot the data beyond the short-term."

Forecast Equation: $\hat{y}_{t+h|t} = l_t + (\phi + \phi^2 + \cdots + \phi^h) b_t$
Level Equation: $l_t = \alpha\, y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1})$
Trend Equation: $b_t = \beta^*\,(l_{t-1} - l_t) + (1-\beta^*)\phi b_{t-1}$

The dampening parameter, $\phi<1$, controls how quickly the trend disappears over time. Predictions are between SES and Holt. $\phi=1$ is Holt's method; $\phi=0$ is SES.
In R, use `holt(, damped=TRUE)`. In SAS use `PROC ESM` with `MODEL=DAMPTREND`.

IV. Exponential Smoothing with Multiplicative Error

Forecast Equation: $\hat{y}_{t+h|t} = l_t\, b_t^h$

Level Equation: $l_t = \alpha\, y_t + (1-\alpha) l_{t-1} b_{t-1}$

Trend Equation: $b_t = \beta^* \left(\frac{l_t}{l_{t-1}}\right) + (1-\beta^*)\, b_{t-1}$

b_t is now a fractional change in level.

It seems that SAS does not implement multiplicative error except for seasonal models (see below). In R, with package "forecast" use `m=ets(y, model="MNN")` for no trend and `m=ets(y, model="MMN", damped=FALSE)` or `m=ets(y, model="MMN", damped=TRUE)` for a trend; follow-up with `forecast(m, h=)` for some specified forecast horizon.

V. Holt-Winters exponential smoothing with seasonality

The seasonal "figure" in a seasonal time series should repeat for each period (m). The Holt-Winters method is more flexible, in that it allows the shape of the figure (but not the period) to vary with time. E.g., each January can have its level (or level and trend) be determined by previous data and α (or α and β) as well as the deviation of January from the zero mean of the whole figure. A new γ parameter is used to model the January deviation as an exponentially declining weighted mean of previous January deviations.

Forecast Equation: $\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m}$

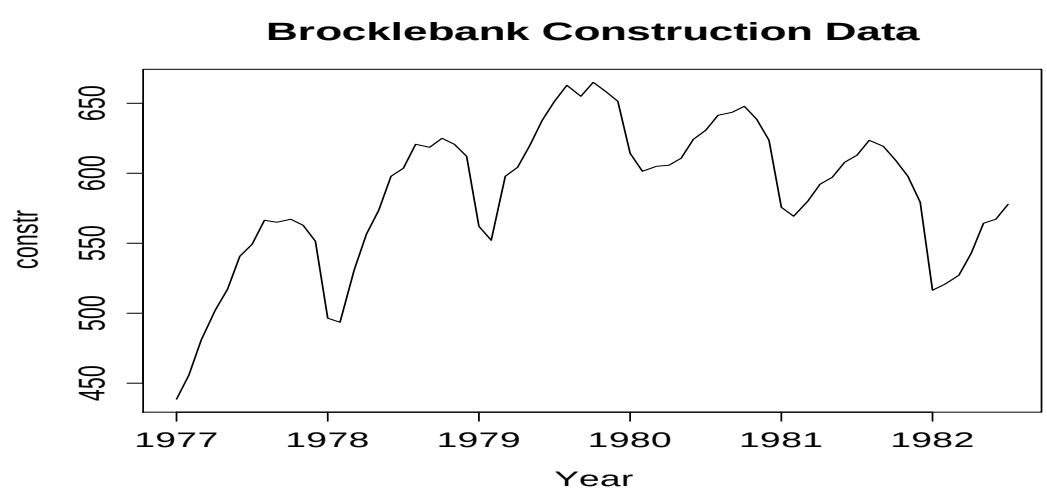
Level Equation: $l_t = \alpha\,(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$

Trend Equation: $b_t = \beta^*\,(l_t - l_{t-1}) + (1-\beta^*) b_{t-1}$

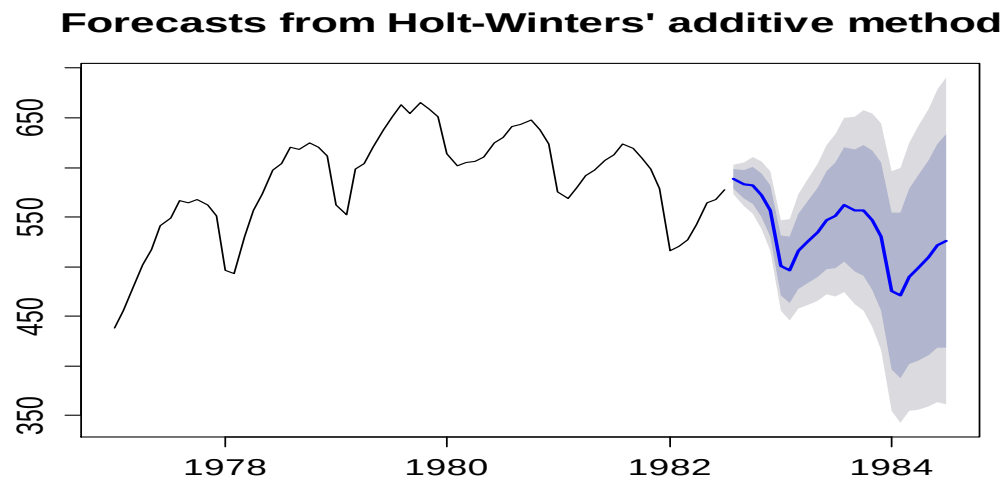
Season Equation: $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma) s_{t-m}$

a. Holt-Winters in R

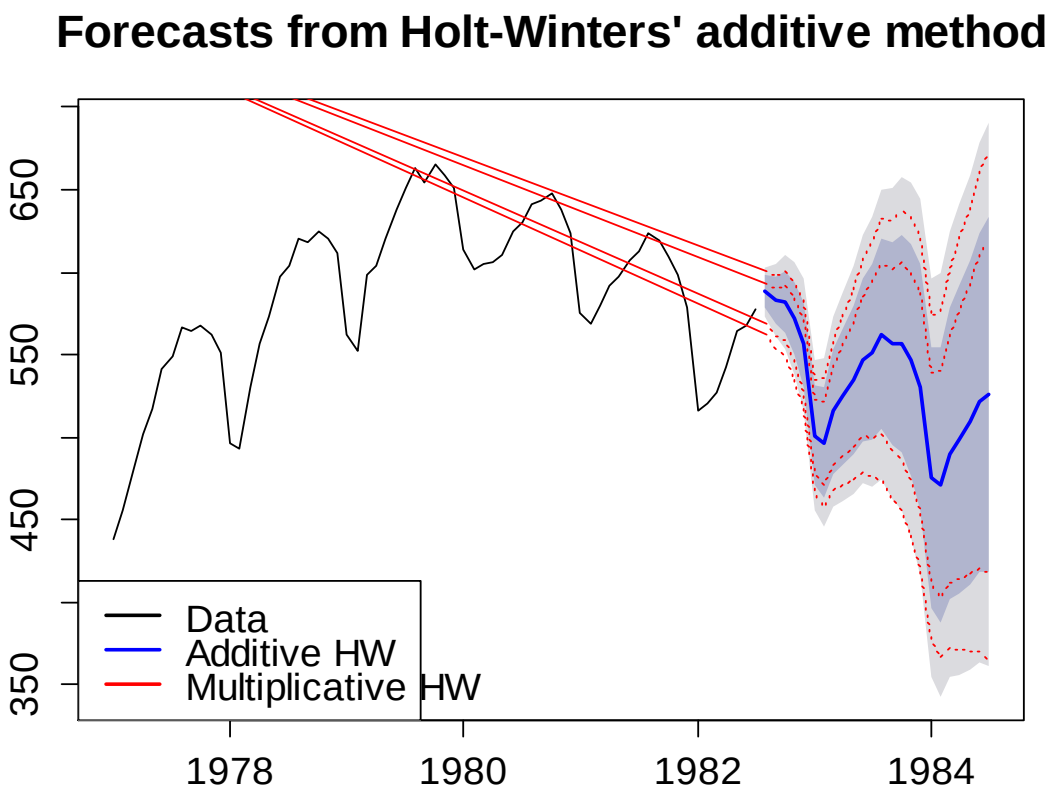
```
constr5 = read.csv("data/const.csv") # from Brocklbank
constr = ts(constr5$CONSTRCT, start=1977, frequency=12)
plot(constr)
```



```
mhw = hw(constr, seasonal="additive", h=24)
plot(mhw)
```



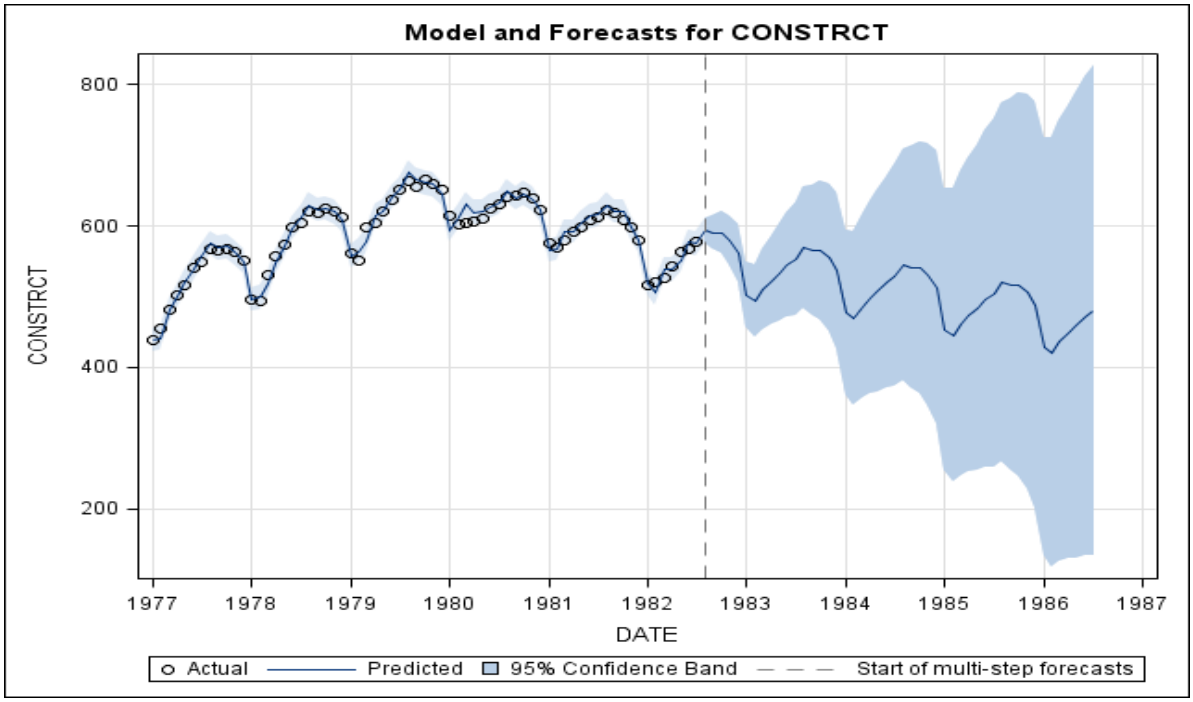

```
summary(mhw)
# Model Information:
# ETS(A,A,A)
# Smoothing parameters: alpha = 0.8895, beta  = 0.1021,
#                       gamma = 2e-04
#   Initial states:
#     l = 491.6553
#     b = 6.4977
#     s=6.9667 20.7899 28.4864 27.066 30.3487 17.3381
#            10.5274 -3.8556 -16.2761 -27.2276 -48.2905 -45.8734
#
#   sigma:  7.5322
#      AIC     AICc      BIC
# 584.2849 595.1649 619.5600
#
# Error measures:
#                     ME     RMSE      MAE        MPE
# Training set -1.262859 7.532167 5.667299 -0.2169746

mhwM = hw(constr, seasonal="multiplicative", h=24)
summary(mhwM)
#   Smoothing parameters: alpha = 0.5584, beta  = 0.0957,
#                         gamma = 1e-04
xplot = function(y, ts.x, ...) {tsp=tsp(ts.x);
lines(seq(tsp[1],tsp[2],by=1/tsp[3]), y, ...)} # get x from ts
xplot(mhwM$lower[,1], mhwM$mean, col=2, lty=3)
xplot(mhwM$lower[,2], mhwM$mean, col=2, lty=3)
xplot(mhwM$upper[,1], mhwM$mean, col=2, lty=3)
xplot(mhwM$upper[,2], mhwM$mean, col=2, lty=3)
legend("bottomleft", c("Data", "Additive HW",
                       "Multiplicative HW"), lwd=2, col=c(1,4,2))
```



Add damped=TRUE?

b. Holt-Winters in SAS

```
%INCLUDE "data/constr.sas";
TITLE1 "Brocklebank Construction Data";
TITLE2 "Holt-Winters Exponential Smoothing";
PROC ESM DATA=const OUT=constPred PRINT=(ESTIMATES FORECASTS)
     LEAD=48 PLOT=(MODELFORECASTS);
  ID date interval = MONTH;
  FORECAST constrct / MODEL=ADDWINTERS;
  *ODS OUTPUT FORECASTS=fcConst; /* if needed */
RUN;
```



Use MODEL=WINTERS for multiplicative error.

VI. Full usage of ets(y, model=, damped=NULL) from package "forecast"

ETS can also stand for error, trend, and seasonality as an alternative to "exponential time series". The 'model' parameter is a string of three letters. Error can be "A", "M", or "Z", for additive, multiplicative, or automatic. Trend and seasonality can be "N" for none, or "A", "M", or "Z". The default is "ZZZ". The default of NULL for 'damped' corresponds to automatic. Parameters are estimated by maximum likelihood (by default), and models are compared by AICc (by default). Not all models are possible. The result can be fed into forecast(object, h=) to forecast for a horizon.