STAT 436 / 536 - Lecture 6

September 19, 2018

Forecasting Strategies - Exponential Smoothing

Often with time series data, our objective is to make future predictions of a time series. Formally we are interested in a response at time $t + k$, x_{t+k}, given that we have observed the time series up to time point t, $\{x_1, \ldots, x_t\}$.

- Assume our process contains no systematic trend or seasonal effects, or that they have been removed.

- *We let the mean of the process change from one time point to the next, but without specifying the direction of the changes.* Formally,

$$x_t = \mu_t + \omega_t,$$

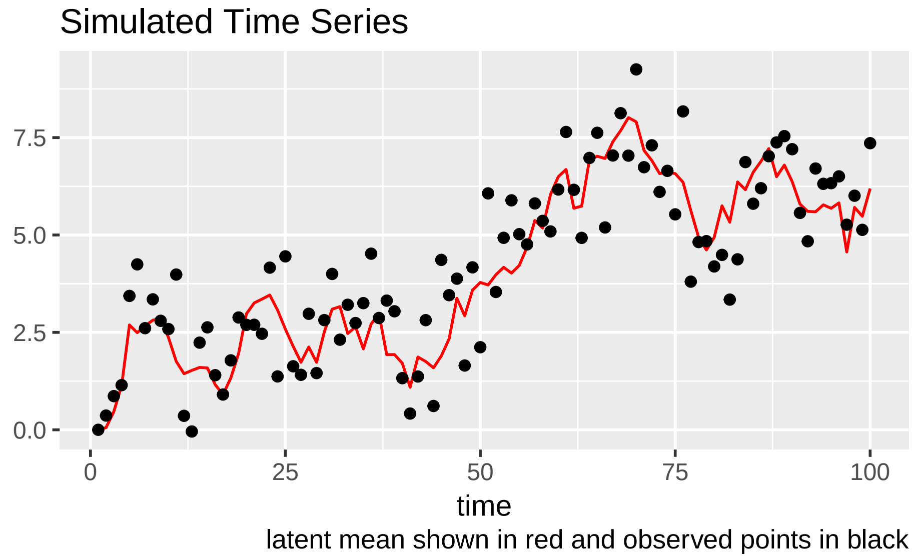
where μ_t is a non-stationary mean and ω_t are independent random deviations, often distributed as $N(0, \sigma^2)$.

- This type of model can be displayed

```
set.seed(09142018)
mu.evol.sd <- .5
sigma <- 1
time.pts <- 100
mu <- rep(0,time.pts)
x <- rep(0, time.pts)

for (t in 2:time.pts){
  mu[t] <- mu[t-1] + rnorm(1,mean=0, sd=mu.evol.sd)
  x[t] <- mu[t] + rnorm(1, mean=0, sd=sigma)
}

ts.df <- data.frame(mu=mu, x=x, time = 1:time.pts)
library(ggplot2)
ggplot(data=ts.df) + geom_line(aes(x = time, y = mu), color = "red") + geom_point(aes(x=time, y=x)) + yl
```



- Explore the impact of changing `mu.evol.sd` and `sigma` in the code above. What are the impacts of these two terms?

- How might these two impact our idea on the next prediction?

Forecasting

- *Assuming that no trend is present, consider the following estimate for* μ

$$\hat{\mu}_t = \alpha x_t + (1-\alpha)\hat{\mu}_{t-1}$$

  - Interpret this estimator, what do we make of α?

  - Does this estimator seem reasonable, yes or no?

  - How would α be influenced by the ratio of `mu.evol.sd` and `sigma`?

- Given that there are no seasonal effects or trends, what is the prediction for $\hat{x}_{t+k} = m\hat{u}_{t+k}$?

- We can rewrite the model in a recursive manner, so that:

$$\hat{\mu}_t = \alpha(x_t - \hat{\mu}_{t-1}) + \hat{\mu}_{t_t} - 1$$

  and then

$$\hat{\mu}_t = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2}\ldots$$

- Typically $\hat{\mu}_1 = x_1$.

- The $\hat{\mu}_t$ is the exponentially weighted moving average and α is the smoothing parameter.

- So we still need to select α, how should we do this?
  - what would be the optimal value of α if $\mu_{t+1} = \mu_t \forall t$ 1/t

  - what about if the change in μ from time to time is much larger than the variance associated with ω_t? Then α should be closer to 1

- The book suggest a default value of $\alpha = .2$ is reasonable, but also defines a method for estimating α.

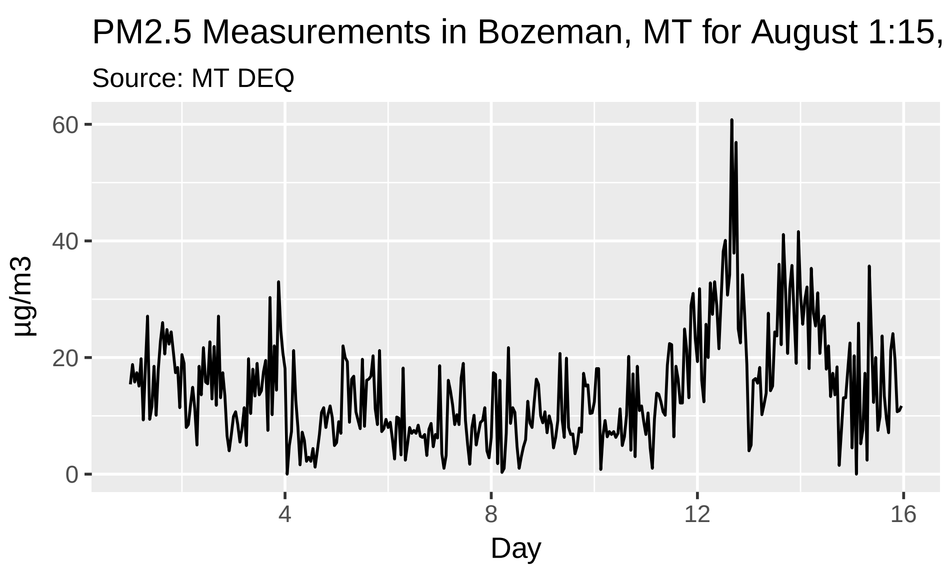
- Let e_t be the one step ahead prediction error, $e_t = x_t - \hat{x}_t = x_t - \hat{x}_t$. Then α can be estimated by minimizing the sum of squared one step ahead prediction error (SS1PE) in a similar fashion to a regression estimate.

- This exponential weighted moving average *is a special case of the Holt-Winters method.*

Air Quality

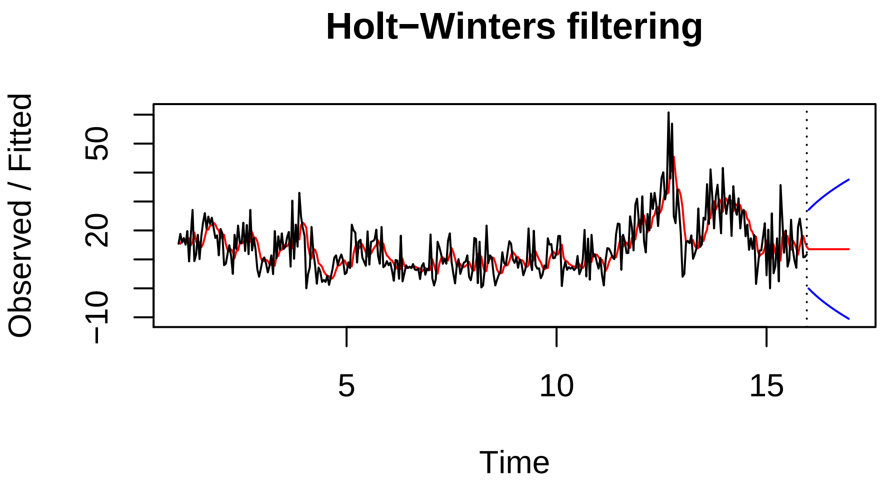
- Recall the air quality measurements taken in Bozeman

```r
air.quality <- scrape_BZNPM_AUG(1:15)
aq <- ts(air.quality$conc, freq = 24, start = 1)
library(ggfortify)
autoplot(aq) + ggtitle(label = 'PM2.5 Measurements in Bozeman, MT for August 1:15, 2018',
                       subtitle = 'Source: MT DEQ') + ylab('µg/m3') + xlab('Day')
```



- There is not clear evidence of cycles or trends in this data set, so we will fit a model without them.

```
aq.hw <- HoltWinters(aq, beta=FALSE, gamma = FALSE, seasonal = 'additive')
alpha.est <- aq.hw$alpha
aq.pred <- predict(aq.hw, n.ahead = 24, prediction.interval = T)
plot(aq.hw, aq.pred)
```



- The α term is estimated to be 0.305.

Holt-Winters Method

- More generally the Holt-Winters method can use exponentially weighted averages to update the mean, or level (which we have already seen); the slope; and the seasonal components.

- Furthermore, these models can be expressed from an additive or multiplicative perspective.

- To express this in an additive framework (3.21 in text) consider

$$\begin{aligned} a_t &= \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) && (1) \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} && (2) \\ s_t &= \gamma(x_t - a_t) + (1-\gamma)s_{t-p} && (3) \end{aligned}$$

where a_t, b_t, and s_t are the estimated level, slope, and seasonal components. The associated smoothing parameters are α, β, and γ. In the previous example a_t was defined as $\hat{\mu}_t$.

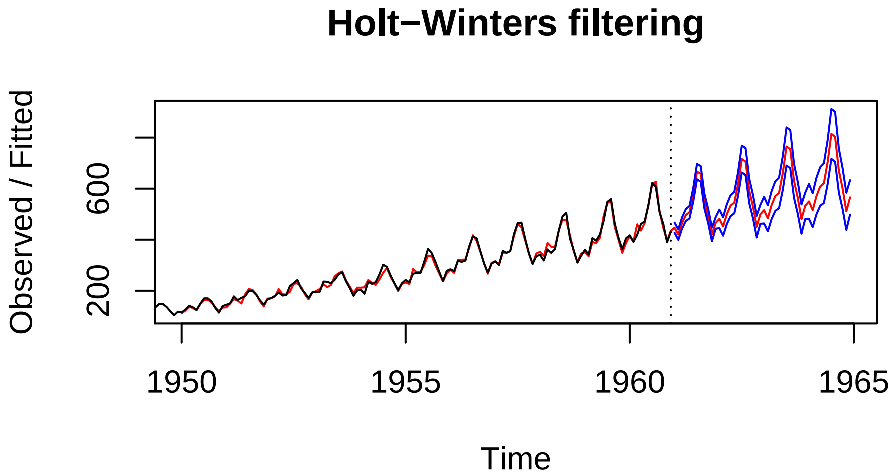
- In this framework the forecasting equation can be written as

$$\hat{x}_{n+k|n} = a_n + kb_n + s_{n+k-p}$$

where we are making predictions at time n for time $n+k$ and p is the length of the seasonal cycle.

- Consider the airline passenger data and a multiplicative decomposition.

```
data("AirPassengers")
AP.hw <- HoltWinters(AirPassengers, seasonal = 'mult')
AP.pred <- predict(AP.hw, n.ahead = 48, prediction.interval = T)
plot(AP.hw,AP.pred)
```



- How do we feel about the prediction here?

- What about if this model was used to predict all the way to today?