Lecture 7: Exponential Smoothing with Trend and Seasonality

And a Brief Tour of Other Popular Forecasting Methods

Introduction to Time Series, Fall 2023

Ryan Tibshirani

Related reading: Chapters 8, 9.10, 12.2, and 12.4 of Hyndman and Athanasopoulos (HA).

1 Simple exponential smoothing

- Exponential smoothing is arguably the other—outside of ARIMA—most popular basic framework for forecasting in time series. These two frameworks bear a neat connection, which you saw at the end of the last lecture on ARIMA, and which we'll revisit a bit later in this lecture

- We'll begin with the simplest possible exponential smoother, called (unsurprisingly?) *simple exponential smoothing* (SES). This constructs a 1-step ahead forecast via

$$\hat{x}_{t+1|t} = \alpha x_t + (1-\alpha)\hat{x}_{t|t-1} \tag{1}$$

  where $\alpha \in [0,1]$ is a parameter to be estimated

- In other words, the SES forecast (1) is a weighted combination of the current observation x_t and the previous forecast $\hat{x}_{t|t-1}$

- By unraveling the iteration, which is basically the same calculation that we did in the ARIMA lecture (but now in the opposite direction), this can also be written as

$$\hat{x}_{t+1|t} = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \dots \tag{2}$$

  This explains its name, since observations x_{t-k} that are k steps into the past are exponentially-downweighted, with weight $(1-\alpha)^k$

- (Note: we are being intentionally vague here about the boundary condition. In ARIMA, to develop the theory cleanly, we let time extend back to $-\infty$. In exponential smoothing, we instead usually index time starting at $t=0$, in which case the right-hand side in (2) would end with $\alpha(1-\alpha)^t x_0$)

- To make h-step ahead forecasts, we iterate (1), where (as usual) we replace future observations by their forecasts. This simply yields $\hat{x}_{t+2|t} = \alpha\hat{x}_{t+1|t} + (1-\alpha)\hat{x}_{t+1|t} = \hat{x}_{t+1|t}$, and in general,

$$\hat{x}_{t+h|t} = \alpha x_t + (1-\alpha)\hat{x}_{t|t-1} \tag{3}$$

  for all horizons $h \geq 1$. That is, SES generates *flat* forecast trajectories. We'll see how to extend this to accomodate a trend, shortly

- While SES smoothing is already very intuitive, we can motivate it in different way, as follows. The *naive flatline forecaster* produces forecasts via

$$\hat{x}_{t+h|t} = x_t \tag{4}$$

  i.e., it just propogates the last observation forward. Meanwhile, the *naive average forecaster* produces forecasts via

$$\hat{x}_{t+h|t} = \frac{1}{t}\sum_{i=1}^{t} x_i \tag{5}$$

  Often we want something in between these two extremes, and that something is given to us by exponential smoothing, recalling the form in (2)

1.1 Component form

- For the developments that follow, it is helpful to rewrite the SES forecast (3) in what is known as *component form*
- Specifically, we think of a "hidden" level ℓ_t that we are tracking over time, that base our forecasts on:

$$\begin{aligned} \hat{x}_{t+h|t} &= \ell_t \\ \ell_t &= \alpha x_t + (1-\alpha)\ell_{t-1} \end{aligned} \tag{6}$$

- The representation in (6) may appear as kind of a trivial rewriting of (3), where we replace $\hat{x}_{t+1|t}$ by ℓ_t, and $\hat{x}_{t|t-1}$ by ℓ_{t-1}. Nonetheless, it serve as a useful jumping off point to extend the model in the next section
- Before moving on, we give a brief example of SES from HA, to forecast internet useage per minute. The data and SES forecast are shown in Figure 1, top row. In order to carry out the forecast, we have to estimate the smoothing parameter α in (6). This is typically done by maximum likelihood (more later), and is what is implemented as the default in the `ETS()` function in the `fable` package
- The forecast from SES is not very impressive, and honestly, in general, SES should probably only be viewed as a small step up from the naive forecasters (4), (5)
- The forecast trajectory from SES is flat, by construction (as previously noted). Next we'll see how to extend the method to accommodate a linear trend

2 Trend extensions

- An extension of the SES forecaster in (6) is *Holt's linear trend* method. This changes the forecast equation (first line) and the level equation (second line) to accomodate an estimate of the slope b_t of the series at time t. And it adds a trend equation to evolve the slope
- Precisely, in component form (which we will stick to henceforth):

$$\begin{aligned} \hat{x}_{t+h|t} &= \ell_t + b_t h \\ \ell_t &= \alpha x_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1} \end{aligned} \tag{7}$$

  Now β is an additional parameter to be estimated, where both $\alpha, \beta \in [0, 1]$
- As before, the level equation updates ℓ_t as an α-weighted combination of the current observation x_t and the previous 1-step ahead forecast $\hat{x}_{t|t-1} = \ell_{t-1} + b_{t-1}$
- The trend equation updates b_t as a β-weighted combination of the current trend $\ell_t - \ell_{t-1}$ and the previous trend b_{t-1}
- Critically, the forecast trajectory from Holt's linear trend method is no longer flat but (as the name suggests, and as is apparent from (7)) a linear function, with slope b_t
- The middle row of Figure 1 shows the forecast from Holt's linear trend method on the internet useage today. To be clear, now both α, β have been estimated from the data. We can see that it predicts a downward trajectory, since b_t at the last time t appears to be negative. However, its prediction intervals are very wide, suggesting that the model is highly uncertain of trend directionality

2.1 Damped trends

- Linear trend forecasts at long horizons can be somewhat erratic; we've already seen that the forecast variance is quite high in the example in Figure 1 (as evidenced by the wide prediction intervals) and this wasn't even a super long horizon ...

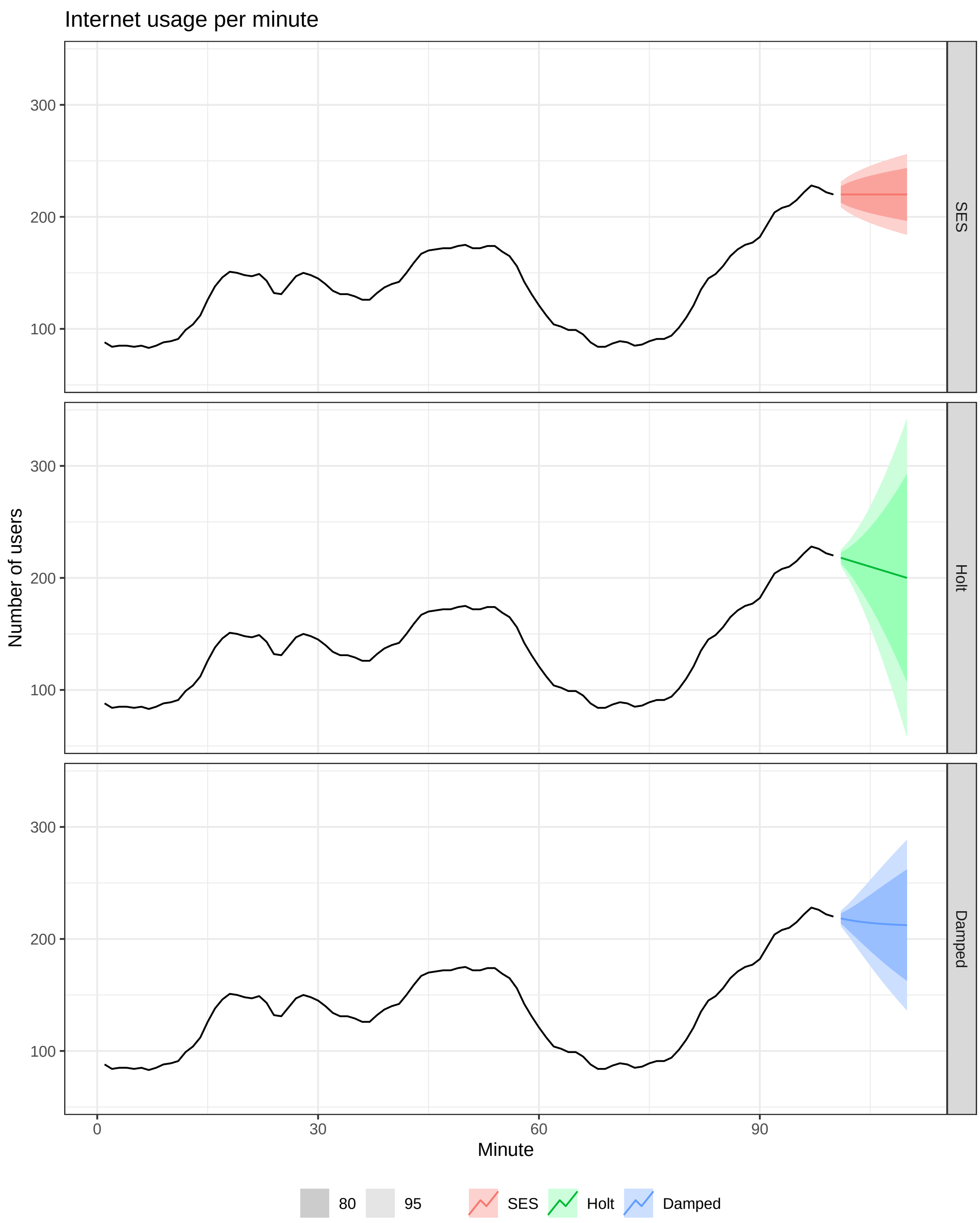

Figure 1: Forecasts on internet useage data (from HA) at 10-steps ahead, from three different exponential smoothing models: simple exponential smoothing (top), Holt's linear trend (middle), and damped linear trend (bottom).

- As a kind of regularization, we can *dampen* the forecasts from Holt's linear trend method. This is often called *damped Holt's* or the *damped linear trend* method

- In addition to $\alpha, \beta \in [0,1]$ as in (7), we introduce a third parameter $\phi \in [0,1]$, to dampen the forecast trajectory:

$$\begin{aligned} \hat{x}_{t+h|t} &= \ell_t + b_t(\phi + \phi^2 + \cdots + \phi^h) \\ \ell_t &= \alpha x_t + (1-\alpha)(\ell_{t-1} + b_{t-1}\phi) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1}\phi \end{aligned} \tag{8}$$

- The interpretation in the damped method (8) is essentially the same as in Holt's method (7), but you can think of it like this: the contribution of a given slope to a forecast diminishes at each step by a multiplicative factor of ϕ

- As $h \to \infty$, the forecasts from the damped linear trend method approach a particular constant (finite) level, namely

$$\hat{x}_{t+h|t} \to \ell_t + b_t \sum_{j=1}^{\infty} \phi^j = \ell_t + b_t \frac{\phi}{1-\phi}$$

  For example, when $\phi = 0.9$, this limit is $\ell_t + 9b_t$

- HA say that a practical range for ϕ is usually 0.8 to 0.98; when ϕ is below 0.8, the dampening is too strong (and short-term forecasts are not "trended" enough); when ϕ is above 0.98, it is too weak (and you cannot distinguish the forecasts from Holt's linear trend method for reasonably horizons). In fact, the `ETS()` function limits the range of ϕ to be $[0.8, 0.98]$, by default

- The bottom row of Figure 1 shows the forecasts from the damped linear trend method on the internet useage data. To be clear, all of α, β, ϕ are estimated from the data. We can see a weak downward trend, that quickly flattens out. The prediction intervals are also much narrower. Inspection of the fitted model (see the R notebook) shows that we the dampening coeffcient estimate is $\hat{\phi} = 0.81$, so it has quite a pronounced effect here

3 Seasonality extensions

- Holt-Winters
- dampening is also possible

3.1 Multiplicative version

-
- dampening is also possible

4 ETS models

- Multiplicative errors and trends
- HA do not recommend multiplicative trend?

4.1 State space representation

-

4.2 Estimation and selection

-

4.3 Forecasting

-

4.4 ARIMA versus ETS

-

5 Theta model

6 Prophet model

7 Neural network AR