Exponential Smoothing

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

Exponential Smoothing

- Exponential smoothing is a standard way to produce forecasts
- We will focus on linear specification
- Large range specifications including multiplicative models
- Linear with logs is a simple way to have a multiplicative model
- Observed data are $\{X_1, X_2, \ldots, X_T\}$

References

1. Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on April 15, 2020.

Simple Exponential Smoothing

Random Walk

- The basic EWMA, which is known as Simple Exponential Smoothing (SES)
- Defined using the recursion

$$\hat{X}_{T+1|T} = \alpha X_T + (1-\alpha)\,\hat{X}_{T|T-1}$$

- The first forecast is

$$\hat{X}_{2|1} = \alpha X_1 + (1-\alpha)\,\hat{X}_{1|0}$$

- $\hat{X}_{1|0}$ is an initial value that is not from the data.
  - Often set to X_1
- α (and $X_{1|0}$) is commonly estimated using least squares

Underlying Model

- Optimal forecast for an Integrated MA(1)

$$X_{T+1} = X_T - (1-\alpha)\,\epsilon_T + \epsilon_{T+1}$$

$$\begin{aligned}
\mathrm{E}_T\left[X_{T+1}\right] &= X_T - (1-\alpha)\,\epsilon_T \\
\epsilon_T &= X_T - X_{T-1} + (1-\alpha)\,\epsilon_{T-1} \\
\Rightarrow \mathrm{E}_T\left[X_{T+1}\right] &= X_T - (1-\alpha)\,X_T + (1-\alpha)\,X_{T-1} - (1-\alpha)^2\,\epsilon_{T-1} \\
&= \alpha X_T + (1-\alpha)\,X_{T-1} - (1-\alpha)^2\,\epsilon_{T-1} \\
&= \alpha X_T + (1-\alpha)\,X_{T-1} - (1-\alpha)^2\,X_{T-1} + (1-\alpha)^2\,X_{T-2} - (1-\alpha)^3\,\epsilon_{T-2} \\
&= \alpha X_T + \alpha\,(1-\alpha)\,X_{T-1} + (1-\alpha)^2\,X_{T-2} + (1-\alpha)^3\,\epsilon_{T-2}
\end{aligned}$$

- If we assume $\epsilon_0 = 0$

$$\mathrm{E}_T\left[X_{T+1}\right] = \alpha \sum_{i=0}^{T} (1-\alpha)^i\,X_{T-i}$$

Forecasts and Prediction Intervals

- Forecasts are always a random walk

$$\hat{X}_{T+h|T} = \hat{X}_{T+1|T}$$

- Prediction Intervals are simple using the IMA analogy

$$PI = \left[\hat{X}_{T+h|T} - 1.96\sigma\sqrt{1+(h-1)\,\alpha^2}, \hat{X}_{T+h|T} + 1.96\sigma\sqrt{1+(h-1)\,\alpha^2}\right]$$

Understanding Prediction Errors

- Note that 2-step forecast error is

$$\begin{aligned} X_{T+2} - \mathrm{E}_T\left[X_{T+2}\right] &= X_{T+2} - \mathrm{E}_T\left[X_{T+1}\right] \\ &= X_{T+2} - \left(X_T - (1-\alpha)\,\epsilon_T\right) \\ &= X_{T+1} - (1-\alpha)\,\epsilon_{T+1} + \epsilon_{T+2} - \left(X_T - (1-\alpha)\,\epsilon_T\right) \\ &= \left(X_T - (1-\alpha)\,\epsilon_T + \epsilon_{T+1}\right) \\ &\qquad - (1-\alpha)\,\epsilon_{T+1} + \epsilon_{T+2} - \left(X_T - (1-\alpha)\,\epsilon_T\right) \\ &= \epsilon_{T+1} - (1-\alpha)\,\epsilon_{T+1} + \epsilon_{T+2} \\ &= \alpha\epsilon_{T+1} + \epsilon_{T+2} \end{aligned}$$

- Same structure continues for any horizon, so

$$X_{T+h} - \mathrm{E}_T\left[X_{T+h}\right] = \epsilon_{T+h} + \sum_{i=1}^{h-1} \alpha\epsilon_{T+i}$$

Conclusions

- Simple Exponential Smoothing is an EWMA
- Forecasts are identical for all horizons
- Depends on two parameters
  - α: the smoothing parameter
  - $X_{1|0}$: the initial value
  - Estimate one or both using LS
- Optimal forecast for an Integrated MA(1)

Exponential Smoothing: Trends and Seasonality

Kevin Sheppard

Advanced Financial Econometrics: Forecasting

May 2020

Holt's Model

Random Walk with Drift

- EWMAs extent to include trends using a two-equation system

$$\begin{aligned} \hat{X}_{T+h|T} &= L_T + hB_T \\ L_T &= \alpha X_T + (1-\alpha)\left(L_{T-1} + B_{T-1}\right) \\ B_T &= \beta\left(L_T - L_{T-1}\right) + (1-\beta)\, B_{T-1} \end{aligned}$$

- Can recursively substitute to gain some insights

$$\begin{aligned} X_{T+1|T} &= L_T + B_T \\ &= \alpha X_T + (1-\alpha)\left(L_{T-1} + B_{T-1}\right) + B_T \\ &= \alpha X_T + (1-\alpha)\left(\alpha X_{T-1} + (1-\alpha)\left(L_{T-2} + B_{T-2}\right)\right) + B_T + (1-\alpha)\, B_{T-1} \\ &= \alpha X_T + \alpha(1-\alpha)\, X_{T-1} + (1-\alpha)^2 L_{T-2} + B_T + (1-\alpha)\, B_{T-1} + (1-\alpha)^2 B_{T-2} \\ &\;\vdots \\ &= \alpha \sum_{i=0}^{T-1} (1-\alpha)^i X_{T-i} + \sum_{j=0}^{T-1} (1-\alpha)^j B_{T-j} \end{aligned}$$

Trend Smoother

- Holt's model contains two exponential smoothers
- Each of the B_T terms are exponential smoothers

$$B_T = \beta \underbrace{\left(L_T - L_{T-1}\right)}_{\text{Innovation to levels}} + (1-\beta)\, B_{T-1}$$

- Known as double exponential smoothing
- Parameters are α and β
- Initial values L_0 and B_0 can be estimated
- Typical fixed values are $L_0 = X_1$ and B_1 are $\left(X_n - X_1\right) / \left(n-1\right)$ for some $n \leq T$
  - Common to choose $n = 2$

Damped Trends

Limiting Trend Effects

- Dampening adds another option
- Local trend growth but ultimately finite

$$
\begin{aligned}
\hat{X}_{T+h|T} &= L_T + \left(\phi + \phi^2 + \ldots + \phi^h\right) B_T \\
L_T &= \alpha X_T + (1-\alpha)\left(L_{T-1} + \phi B_{T-1}\right) \\
B_T &= \beta\left(L_T - L_{T-1}\right) + (1-\beta)\,\phi B_{T-1}
\end{aligned}
$$

- Long-run forecast converges to

$$
\lim_{h \to \infty} X_{T+h|T} = L_T + \frac{\phi}{1-\phi} B_T
$$

- ϕ is usually forced to be large
  - Damped trend forecast is not too far from the trend over short horizons
- If $\phi \approx 1$ then difficult to distinguish from undamped model
- Usually restricted to $0.8 \leq \phi \leq 0.98$
- In practice, damping helps

Holt-Winters

Random walk with drift and Seasonality

- Seasonality is added to produce the Holt-Winters model
- Four equation system

$$
\begin{aligned}
\hat{X}_{T+h|T} &= L_T + hB_T + S_{T+h-m(k+1)} \\
L_T &= \alpha\underbrace{(X_T - S_{T-m})}_{\text{Deseasonalized}} + (1-\alpha)\left(L_{T-1} + B_{T-1}\right) \\
B_T &= \beta\left(L_T - L_{T-1}\right) + (1-\beta)\, B_{T-1} \\
S_T &= \gamma\underbrace{(X_T - L_{T-1} - B_{T-1})}_{\text{Surprise rel Level and Trend}} + (1-\gamma)\, S_{T-m}
\end{aligned}
$$

- System of $2+m$ smoothers
- Each seasonality is as its own EWMA/SES using a modified shock
- $k = \lfloor (h-1)/m \rfloor$

Damped Trend

Limiting Trend Effects

- Can be damped to reduce trend effects

$$\hat{X}_{T+h|T} = L_T + \left(\phi + \phi^2 + \ldots + \phi^h\right) B_T + S_{T+h-m(k+1)}$$

$$L_T = \alpha \left(X_T - S_{T-m}\right) + (1-\alpha)\left(L_{T-1} + \phi B_{T-1}\right)$$

$$B_T = \beta \left(L_T - L_{T-1}\right) + (1-\beta)\,\phi B_{T-1}$$

$$S_T = \gamma \left(X_T - L_{T-1} - \phi B_{T-1}\right) + (1-\gamma)\,S_{T-m}$$

- Damping parameter $0.8 \leq \phi \leq 0.98$

Prediction Intervals

- 95% prediction intervals are complicated in the larger models

$$PI = \left[\hat{X}_{T+h|T} \pm 1.96\sigma\psi_h\right]$$

- Trend

$$\psi_h^2 = 1 + (h-1)\left[\alpha^2 + \alpha^2\beta h + \frac{1}{6}\alpha^2\beta^2 h\,(2h-1)\right]$$

- Damped Trend

$$\begin{aligned}\psi_h^2 =& 1 + \alpha^2 (h-1) + \frac{\alpha\beta\phi h}{(1-\phi)^2}\left[2\alpha(1-\phi) + \alpha\beta\phi\right] \\ &- \frac{\alpha\beta\phi\left(1-\phi^h\right)}{(1-\phi)^2\left(1-\phi^2\right)}\left[2\alpha\left(1-\phi^2\right) + \alpha\beta\phi\left(1+2\phi-\phi^h\right)\right]\end{aligned}$$

- Holt-Winters

$$\psi^2 = 1 + (h-1)\left[\alpha^2 + \alpha^2\beta h + \frac{1}{6}\alpha^2\beta^2 h\,(2h-1)\right] + \gamma k\left\{2\alpha + \gamma + \alpha\beta m\,(k+1)\right\}$$

- Damped Holt-Winters: Too long to type!

Conclusions

- Holt's Models includes two exponential smoothers
  - Level and Trend
- The Holt-Winters Model uses $2 + m$ exponential smoothers
  - Level, Trend, and Seasonal (m)
- Damping the trend adds additional flexibility
  - Limits the long-run effect of the trend on the forecast
  - Generally improves performance
- Prediction intervals are available for all four specifications