Exponential Smoothing

BSAD 8310: Business Forecasting — Lecture 3

Department of Economics

University of Nebraska at Omaha

Spring 2026

Motivation: Adaptive Averaging

Regression treats all observations equally — but recent observations carry more information.

In OLS regression, *every* observation receives equal implicit weight:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad \Rightarrow \qquad \hat{y}_{T+h|T} = \mathbf{x}'_{T+h}\hat{\boldsymbol{\beta}}$$

When the data-generating process **drifts over time** — structural breaks, changing growth rates, evolving seasonality — old observations **contaminate** the forecast.

A retailer's sales pattern shifts after a store renovation. OLS keeps fitting to pre-renovation data, systematically over- or under-forecasting for months afterward.

Idea: assign geometrically declining weights to past observations.

Observation	Age	Weight
y_T	0	α
y_{T-1}	1	$\alpha(1-\alpha)$
y_{T-2}	2	$\alpha(1-\alpha)^2$
y_{T-j}	j	$\alpha(1-\alpha)^j$

Weights sum to 1: $\sum_{j=0}^{\infty} \alpha(1-\alpha)^j = 1$ for $0 < \alpha \leq 1$.

A single parameter α controls how fast old observations are **forgotten**. $\alpha \approx 1$: almost all weight on the latest value. $\alpha \approx 0$: long memory.

Simple Exponential Smoothing

One parameter, one component: the level.

Let ℓ_t denote the **level** at time t. SES updates:

$$\ell_t \;=\; \alpha\, y_t \;+\; (1-\alpha)\,\ell_{t-1}, \qquad 0 < \alpha \leq 1$$

with initialization ℓ_0 estimated from the data.

Recursive expansion reveals the exponential weights:

$$\ell_t \;=\; \alpha \sum_{j=0}^{t-1} (1-\alpha)^j\, y_{t-j} \;+\; (1-\alpha)^t\, \ell_0$$

As $t \to \infty$, the initialization term $(1-\alpha)^t \ell_0 \to 0$ for any $\alpha > 0$ — the starting value eventually becomes irrelevant.

SES can be rewritten as an **error-correction equation**:

$$\ell_t = \ell_{t-1} + \alpha \underbrace{(y_t - \ell_{t-1})}_{e_t} = \ell_{t-1} + \alpha\, e_t$$

where $e_t = y_t - \ell_{t-1}$ is the one-step-ahead forecast error. *(Recall from Lecture 2: SES shares the AR(1) error-correction structure, but requires no stationarity.)*

At each period the level is **updated in proportion to the latest error**. α controls the speed of adaptation.

Large α (e.g. 0.8):

- Reacts quickly to shocks
- Noisy, volatile forecasts
- Good for rapidly changing series

Small α (e.g. 0.1):

- Slow adaptation
- Smooth, stable forecasts
- Good for slowly evolving, near-stationary series

Parameter estimation: choose α and ℓ_0 to minimise the sum of squared one-step-ahead errors (Hyndman and Athanasopoulos 2021, Ch. 8):

$$\min_{\alpha,\,\ell_0} \sum_{t=1}^{T} (y_t - \ell_{t-1})^2, \qquad 0 < \alpha \leq 1$$

Solved numerically (bounded nonlinear optimisation).

Forecast: for *any* horizon $h \geq 1$,

$$\hat{y}_{T+h|T} \;=\; \ell_T$$

> SES produces a **flat forecast** — the same value at all horizons. Appropriate only when the series has *no trend and no seasonality*.

Socratic: if $y_T = 110$, $\ell_{T-1} = 104$, and $\hat{\alpha} = 0.6$, what is $\hat{y}_{T+1|T}$?

Estimated $\hat{\alpha} \approx 0.70$ — fast adaptation. Forecast for month $T+1$ equals the final smoothed level ℓ_T. Lab 3 compares RMSE against Lecture 1–2 benchmarks.

Connection to benchmarks:

Special case	**Model**
$\alpha = 1$	Naïve forecast: $\hat{y}_{T+h\|T} = y_T$
$\alpha \to 0$ (large T)	Historical mean: $\hat{y}_{T+h\|T} = \bar{y}$

SES **nests** both extreme benchmarks from Lecture 1. $\hat{\alpha}$ locates the optimal point between "forget everything" and "forget nothing."

Holt's Linear Method

Add a second component for trend: level and slope updated separately.

Level: $\ell_t = \alpha\, y_t + (1-\alpha)\,(\ell_{t-1} + b_{t-1})$

Trend: $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\, b_{t-1}$

Parameters: $0 < \alpha \leq 1,\ \ 0 < \beta^* \leq 1$ (Holt 2004)

Interpretation:

- ℓ_t: where the series is *now*
- b_t: the current growth rate
- β^* controls how fast the slope adapts

SES is nested inside Holt: set $\beta^* = 0$ and $b_0 = 0$.

h-step forecast:

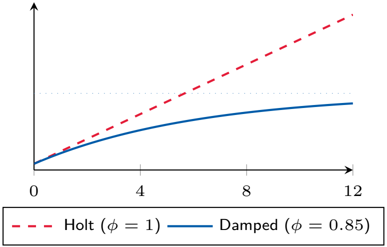
$$\hat{y}_{T+h|T} = \ell_T + h\, b_T$$

Extrapolates the current trend linearly.

Linear extrapolation can be unrealistic for long horizons — trends rarely persist indefinitely.

Damped-trend modification (Gardner 2006):

$$\hat{y}_{T+h|T} = \ell_T + \sum_{j=1}^{h} \phi^j b_T, \quad 0 < \phi < 1$$



$\phi \to 1$: standard Holt (no damping).
$\phi \to 0$: forecast $\to \ell_T$ (flat; trend vanishes).
Use damped by default for $h > 6$.

Socratic: with $\ell_T = 100$, $b_T = 4$, $\phi = 0.85$, what is $\hat{y}_{T+3|T}$? (Hint: compute $\phi + \phi^2 + \phi^3$.) Why might large β^* cause erratic forecasts when spikes occur?

Holt-Winters Seasonal Method

Add a third component for seasonality: level, trend, and season.

The **additive Holt-Winters model** (Winters 1960):

$$y_t = \ell_t + b_t + s_t + \varepsilon_t$$

where s_t is the seasonal component with period m.

$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\, b_{t-1}$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)\, s_{t-m}$$

Parameters: $0 < \alpha \leq 1$, $0 < \beta^* \leq 1$, $0 < \gamma \leq 1-\alpha$; seasonal period m

The m seasonal indices sum to zero over one full period. For monthly data, $m = 12$.

Additive forecast:

$$\hat{y}_{T+h|T} = \ell_T + h\, b_T + s_{T+h-m}$$

Multiplicative forecast:

$$\hat{y}_{T+h|T} = (\ell_T + h\, b_T)\, s_{T+h-m}$$

Variant	Seasonal component	When to use
Additive	$\ell_t + s_t$	Seasonal amplitude *constant* in size
Multiplicative	$\ell_t \times s_t$	Seasonal amplitude *proportional* to level

The holiday-season spike grows as overall sales grow $\Rightarrow$ **multiplicative seasonality** is usually preferred.

Parameters to estimate: α, β^*, γ, plus initial values ℓ_0, b_0, and m seasonal indices $s_{1-m}, \ldots, s_0$.

Initialization strategies:

- **Classical:** first-season average for ℓ_0; first-to-second-season slope for b_0; seasonal differences for $s_1, \ldots, s_m$
- **Optimized:** jointly minimize SSE over all parameters *and* initial values (default in `statsmodels`)

Holt-Winters requires at least $\mathbf{2m}$ observations to initialise. With $m = 12$, you need ≥ 24 months of data.

Socratic: if the December spike triples over ten years, which variant is appropriate and what does that imply for $\hat{\gamma}$?

The ETS Framework

A unified state space taxonomy: Error $\times$ Trend $\times$ Seasonal.

Hyndman and Athanasopoulos (2021) unify all exponential smoothing variants as **state space models** with a single error source.

Component	**Options**	**Meaning**
Error (E)	A, M	Additive / Multiplicative
Trend (T)	N, A, A_d	None / Additive / Additive damped
Seasonal (S)	N, A, M	None / Additive / Multiplicative

$2 \times 3 \times 3 = 18$ combinations; some are inadmissible, giving **15 valid ETS models**.

ETS(A,N,N) $=$ SES
ETS(A,A,N) $=$ Holt linear
ETS(A,$\mathbf{A_d}$,N) $=$ Holt damped

ETS(A,A,A) $=$ Holt-Winters additive
ETS(A,A,M) $=$ Holt-Winters multiplicative
(and 10 more variants)

Every ETS model can be written as a **state space model**:

$$y_t = h(\mathbf{x}_{t-1}) + k(\mathbf{x}_{t-1})\,\varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + g(\mathbf{x}_{t-1})\,\varepsilon_t$$

where $\mathbf{x}_t = (\ell_t,\, b_t,\, s_t, \ldots)'$ is the state vector.

Additive error (`E = A`):

- Gaussian likelihood, closed-form prediction interval (PI)
- Easier to interpret

Multiplicative error (`E = M`):

- Often better RMSE
- No closed-form PI

The state space form enables **maximum likelihood estimation** and principled prediction intervals.

Notation: ε_t *here is the model innovation. For additive-error models,* $\varepsilon_t \equiv e_t$ *(forecast error); for multiplicative-error models,* $\varepsilon_t = e_t/\ell_{t-1}$.

Fit all 15 valid ETS models; select by minimum AIC (Hyndman and Athanasopoulos 2021, Ch. 8):

$$\mathrm{AIC} = -2\hat{\ell} + 2p, \qquad \hat{\ell} = \text{maximized log-likelihood}$$

ETS model	**Key features**	**Typical application**
ETS(A,N,N)	No trend, no seasonal	Stable, level series
ETS(A,A_d,N)	Damped trend	Long-horizon trending
ETS(A,A,M)	Linear trend, mult. seasonal	Retail, tourism

In the M4 Competition, ETS-family methods ranked among the most accurate purely statistical approaches (Makridakis et al. 2020).

Lab 3 uses `statsmodels.tsa.exponential_smoothing.ets.ETSModel` for full AIC-based model selection across all valid ETS variants.

Exponential weighting gives more importance to recent observations; α controls the forgetting rate.

SES: one parameter, flat forecast — nests naïve ($\alpha{=}1$) and mean ($\alpha{\to}0$).

Holt: adds trend component b_t — use the *damped* variant for horizons $h > 6$.

Holt-Winters: adds seasonal component s_t — choose additive vs. multiplicative based on whether amplitude scales with the level.

ETS framework: unifies all methods as state space models; AIC selects the best variant automatically.

All ETS models share the error-correction intuition: update each component in proportion to the latest forecast error.

Exponential smoothing captures **level, trend, and seasonality** through adaptive updating.

What ETS cannot easily handle:

- Rich autocorrelation structure beyond a single AR(1)-like decay
- Integration and differencing in a principled statistical framework
- Combining external regressors with adaptive dynamics (ARIMAX)

Lecture 4: Stationarity, differencing, and the ARIMA model family — a regression-based approach to capturing autocorrelation.

Lab 3: Fit SES, Holt, Holt-Winters, and auto-ETS to RSXFS; compare RMSE / MAE with Lecture 1–2 benchmarks.

- Gardner, Everette S. (2006). "Exponential Smoothing: The State of the Art—Part II". In: *International Journal of Forecasting* 22.4, pp. 637–666.
- Holt, Charles C. (2004). "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages". In: *International Journal of Forecasting* 20.1. Reprint of the 1957 ONR Research Memorandum, pp. 5–10.
- Hyndman, Rob J. and George Athanasopoulos (2021). *Forecasting: Principles and Practice*. 3rd. Open access. Melbourne, Australia: OTexts. URL: https://otexts.com/fpp3/.
- Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos (2020). "The M4 Competition: 100,000 Time Series and 61 Forecasting Methods". In: *International Journal of Forecasting* 36.1, pp. 54–74.
- Winters, Peter R. (1960). "Forecasting Sales by Exponentially Weighted Moving Averages". In: *Management Science* 6.3, pp. 324–342.