

Time Series Analysis

ETS

Dr. Ludovic Amruthalingam
ludovic.amruthalingam@hslu.ch

Informatik



Outline

- Simple exponential smoothing
- (Damped) Holt's linear trend
- Holt-Winters methods
- Time series decompositions with Holt-Winters
- Innovation state space models – ETS
- Prediction confidence intervals
- Connection to ARIMA

Simple exponential smoothing (SES)

Given a TS realization $\{x_1, x_2, \dots, x_t, \dots, x_n\}$, the 1-step ahead forecast is modeled as a weighted average (avg_w) of the **current observation** x_t and the **previous 1-step ahead forecast**.

$$\hat{x}_{t+1|t} = \alpha x_t + (1 - \alpha) \hat{x}_{t|t-1} \text{ with } \alpha \in [0,1].$$

Choosing a **large (small)** α value assigns **more weight to recent (older) observations**.

Substituting the intermediate forecasts: $\hat{x}_{t+1|t} = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots$

$$\hat{x}_{t+1|t} = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i x_{t-i} + (1 - \alpha)^t l_0 \text{ with } \hat{x}_{1|0} \text{ denoted as } l_0.$$

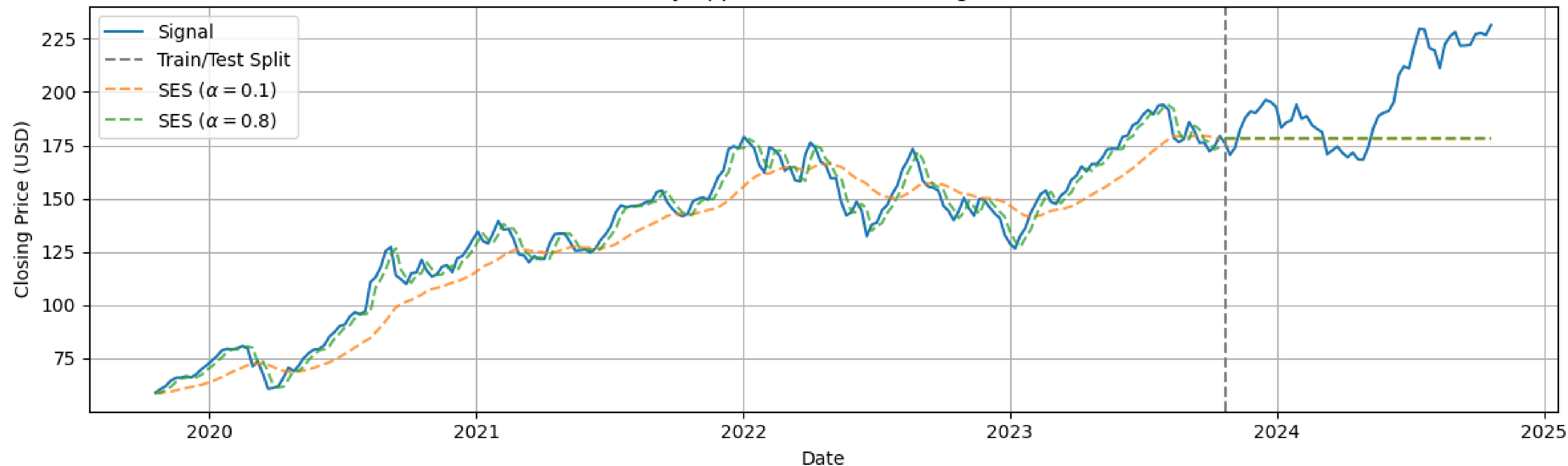
i.e., forecast as weighted averages of past observations with exponentially decaying weights.

SES generates **flat forecast trajectories**

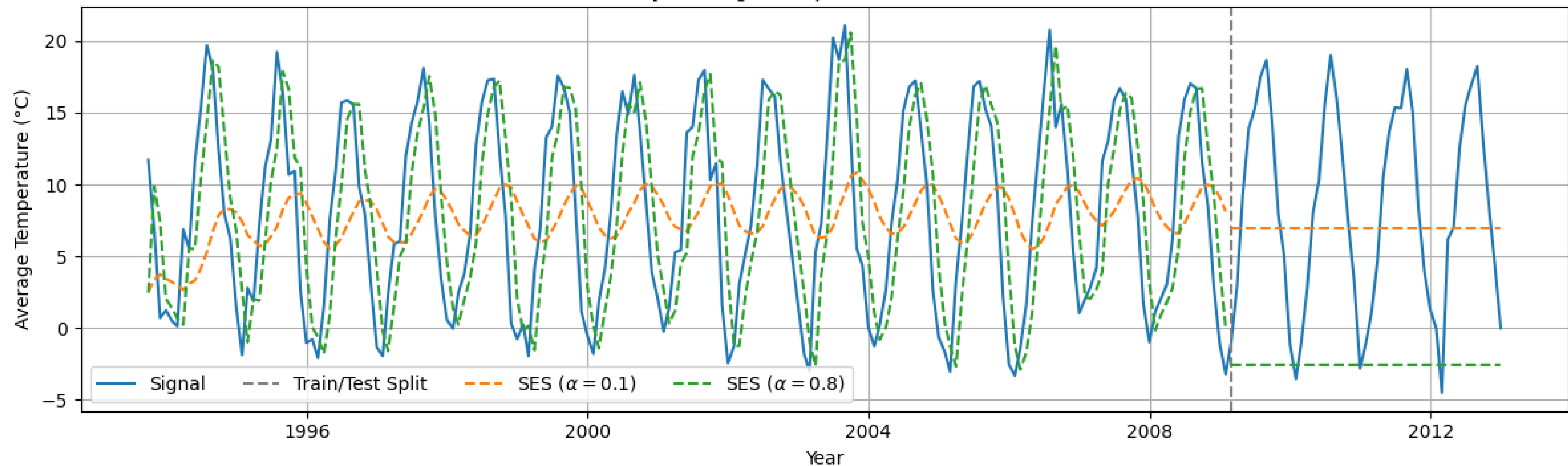
$$\hat{x}_{t+h|t} = \alpha \hat{x}_{t+h-1|t} + (1 - \alpha) \hat{x}_{t+h-1|t} = \hat{x}_{t+h-1|t} = \dots = \hat{x}_{t+1|t}$$

SES assumes the data fluctuate randomly around a **constant mean level: data without trend or seasonality**.

Weekly Apple Inc. (AAPL) Closing Prices



Monthly Average Temperature in Switzerland



SES component form

Represent model components **separately**.

SES only considers the **level** l_t of the time series.

Forecast equation: $\hat{x}_{t+h|t} = l_t$

Level equation: $l_t = \alpha x_t + (1 - \alpha)l_{t-1}$ with $\alpha \in [0,1]$

l_t is the avg_w of observation x_t and the previous level i.e., the previous 1-step ahead forecast $\hat{x}_{t|t-1} = l_{t-1}$.

Estimate α and l_0 from $\{x_1, x_2, \dots, x_T\}$ by minimizing the **sum of squared errors** (SSE)

$$SSE = \sum_{i=1}^T (x_t - \hat{x}_{t|t-1})^2$$

Holt's linear trend method

Extend SES by adding an estimate of the level's **rate of change**, called the **trend component**.

Forecast equation:	$\hat{x}_{t+h t} = l_t + hb_t$	
Level equation:	$l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + b_{t-1})$	
Trend equation:	$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$	with $\alpha, \beta^* \in [0,1]$

l_t is the avg_w of observation x_t and the previous 1-step ahead forecast $\hat{x}_{t|t-1} = l_{t-1} + b_{t-1}$.
 b_t is the avg_w of the current trend estimate $l_t - l_{t-1}$ and the previous trend b_{t-1} .

Generates **linear forecast trajectories** function of h .

Estimate $\alpha, l_0, \beta^*, b_0$ from $\{x_1, x_2, \dots, x_T\}$ by minimizing the SSE.

Damped Holt's linear trend method

Purely linear trend tend to **over-forecast** especially for longer forecast horizons → consider **damped trend**.

Forecast equation: $\hat{x}_{t+h|t} = l_t + b_t \sum_{i=1}^h \phi^i$
Level equation: $l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$
Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1}$ with $\alpha, \beta^*, \phi \in [0,1]$

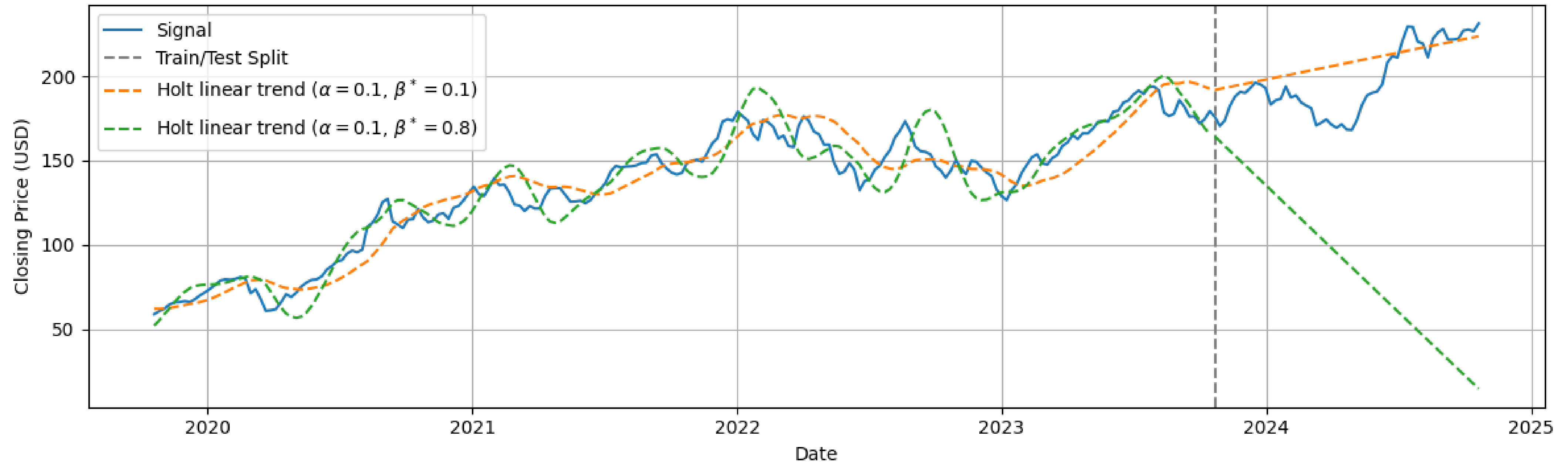
For $\phi = 1$, the method is equivalent to Holt's linear trend method.

The damping effect is **stronger (weaker)** for **smaller (larger)** ϕ values.

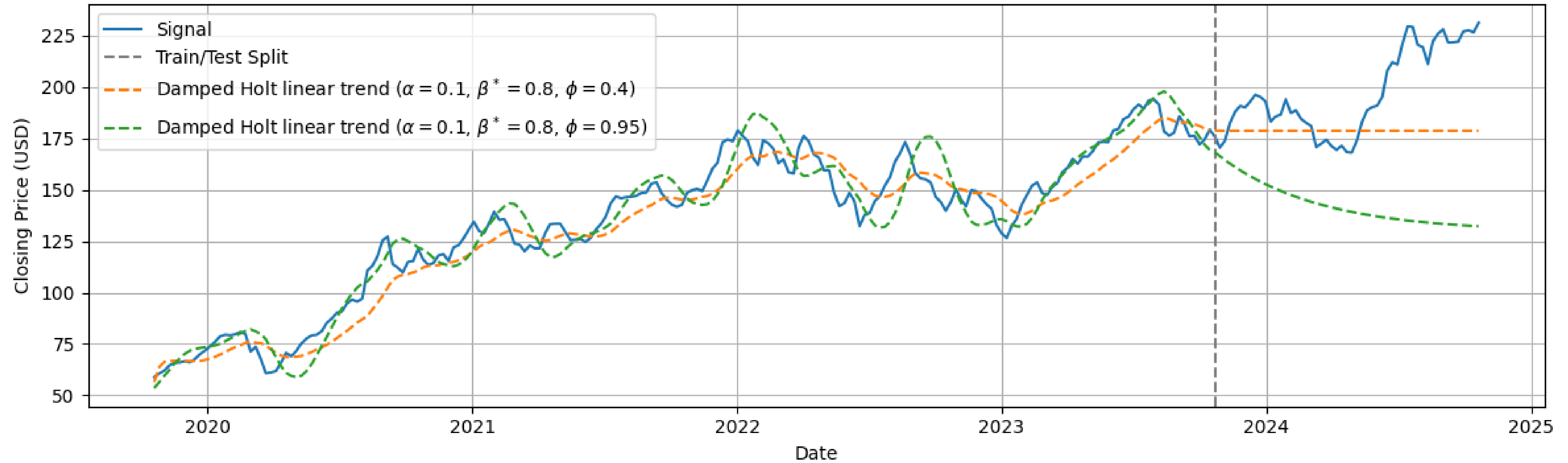
For long horizons (and provided $\phi < 1$), the trend **tends to a constant** since $\lim_{h \rightarrow \infty} \hat{x}_{t+h|t} = l_t + b_t \frac{\phi}{1-\phi}$.

Estimate $\alpha, l_0, \beta^*, b_0, \phi$ from $\{x_1, x_2, \dots, x_T\}$ by minimizing the SSE. In practice ϕ typically lies within $[0.8, 0.98]$.

Weekly Apple Inc. (AAPL) Closing Prices



Weekly Apple Inc. (AAPL) Closing Prices



Holt-Winters additive method

Extend Holt's linear trend method by adding an **estimate of the seasonal component** of the time series.

Forecast equation: $\hat{x}_{t+h|t} = l_t + hb_t + s_{t+h-P[h/P]}$
Level equation: $l_t = \alpha(x_t - s_{t-P}) + (1 - \alpha)(l_{t-1} + b_{t-1})$
Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$
Seasonal equation: $s_t = \gamma^*(x_t - l_t) + (1 - \gamma^*)s_{t-P}$ with $\alpha, \beta^*, \gamma^* \in [0,1]$

l_t is the avg_w of the seasonality adjusted observation $x_t - s_{t-P}$ and the previous 1-step ahead non-seasonal forecast.

b_t is the avg_w of the current trend estimate $l_t - l_{t-1}$ and the previous trend b_{t-1} .

s_t is the avg_w of the level adjusted observation $x_t - l_t$ and the seasonal index from the previous period.

Replacing l_t in the seasonal equation yields $s_t = \gamma(x_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-P}$ with $\gamma = \gamma^*(1 - \alpha)$ and $\gamma \in [0, 1 - \alpha]$.

s_t is the avg_w of the current seasonal index estimate $x_t - l_{t-1} - b_{t-1}$ and the seasonal index from the previous period.

Estimate $\alpha, l_0, \beta^*, b_0, \gamma, s_{0,-1,-2,\dots,-(P-1)}$ from $\{x_1, x_2, \dots, x_T\}$ by minimizing the SSE.

Holt-Winters multiplicative method

Use Holt-Winters' **additive method** when the seasonal variations are roughly **constant**.

- $s_{t+1, \dots, t+P}$ are expressed in absolute terms with the same unit as the TS and sum up to approximately zero.

When the seasonal variations are roughly **proportional** to the TS level, consider either non-linear transforms or the Holt-Winters' **multiplicative method**.

- $s_{t+1, \dots, t+P}$ are expressed in relative terms and sum up to approximately P .

Forecast equation:

$$\hat{x}_{t+h|t} = (l_t + hb_t)s_{t+h-P[h/P]}$$

Level equation:

$$l_t = \alpha \frac{x_t}{s_{t-P}} + (1 - \alpha)(l_{t-1} + b_{t-1})$$

Trend equation:

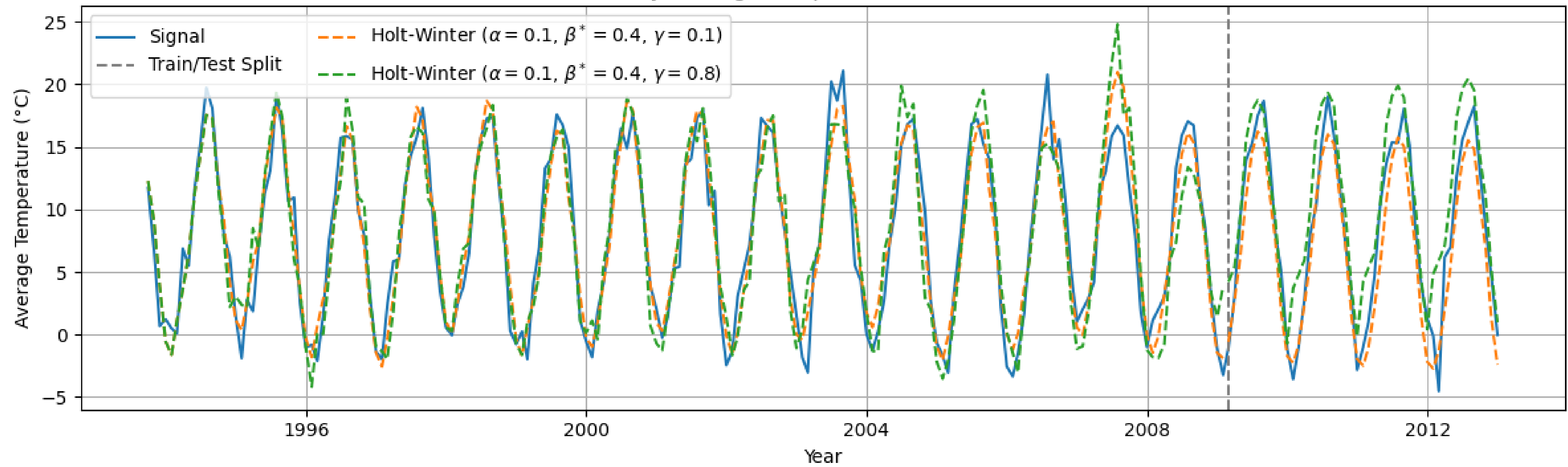
$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$$

Seasonal equation:

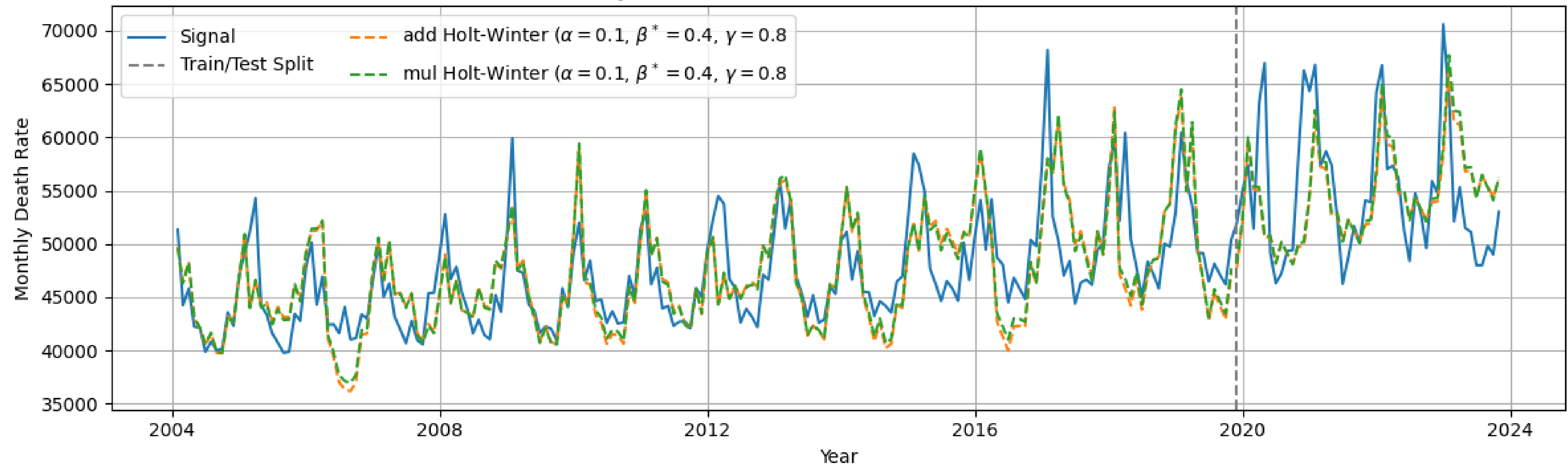
$$s_t = \gamma \frac{x_t}{l_{t-1} + b_{t-1}} + (1 - \gamma)s_{t-P}$$

with $\alpha, \beta^* \in [0,1]$ and $\gamma \in [0, 1 - \alpha]$

Monthly Average Temperature in Switzerland



Monthly Death Rates Over the Last 20 Years in France



Holt-Winters' damped method

Damped trend and additive seasonality

Forecast equation: $\hat{x}_{t+h|t} = l_t + b_t \sum_{i=1}^h \phi^i + s_{t+h-P[h/P]}$
Level equation: $l_t = \alpha(x_t - s_{t-P}) + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$
Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
Seasonal equation: $s_t = \gamma(x_t - l_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-P}$

with $\alpha, \beta^*, \phi \in [0,1]$ and $\gamma \in [0,1 - \alpha]$

Damped trend and multiplicative seasonality

Forecast equation: $\hat{x}_{t+h|t} = (l_t + b_t \sum_{i=1}^h \phi^i) s_{t+h-P[h/P]}$
Level equation: $l_t = \alpha \frac{x_t}{s_{t-P}} + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$
Trend equation: $b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
Seasonal equation: $s_t = \gamma \frac{x_t}{l_{t-1} - \phi b_{t-1}} + (1 - \gamma)s_{t-P}$

with $\alpha, \beta^*, \phi \in [0,1]$ and $\gamma \in [0,1 - \alpha]$

Time series decomposition with Holt-Winters

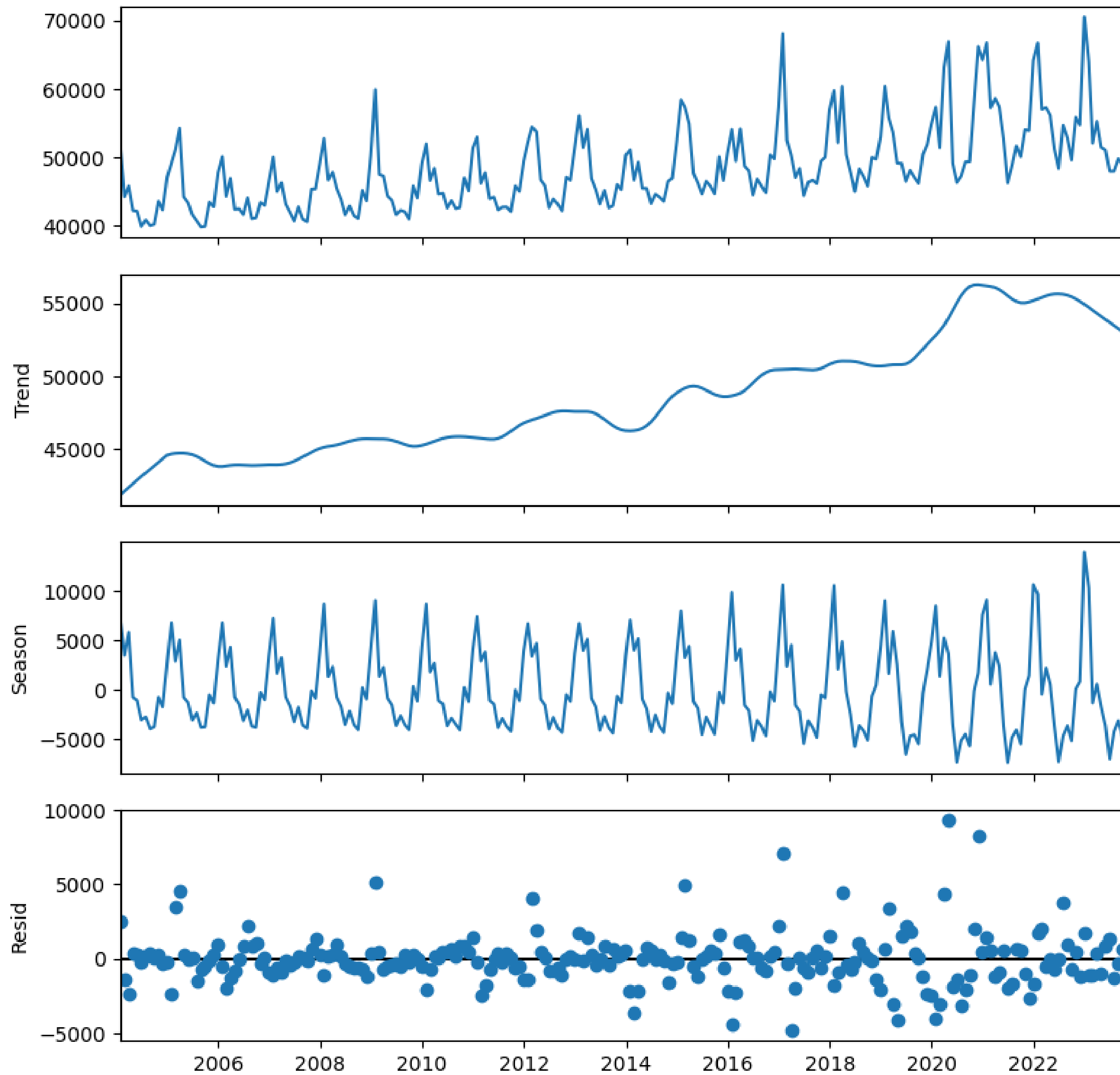
Holt-Winters' components provides a decomposition comparable to decomposition methods like STL decomposition.

Note that Holt-Winters' **level component** corresponds to the **trend component** of decomposition methods.

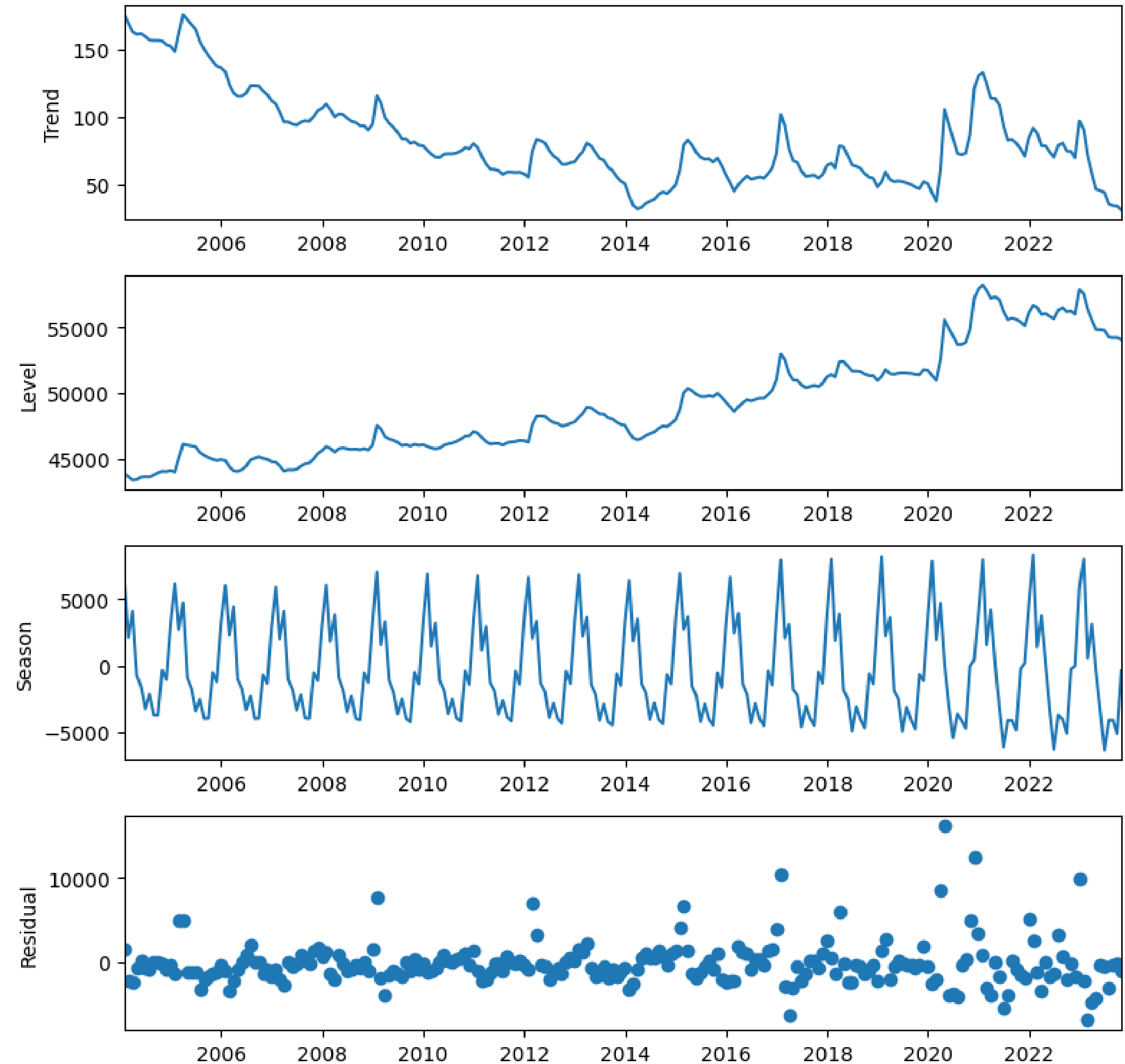
Holt-Winters' **trend component** has **no counterpart** in decomposition methods.

Holt-Winters' trend component can be understood as the projected **direction** or **trajectory** of the time series.

STL decomposition: Monthly Death Rates Over the Last 20 Years in France



Holt-Winter Components: Monthly Death Rates Over the Last 20 Years in France



Innovation state space models – ETS

The methods described so far are **deterministic** and cannot provide **uncertainty estimates**.

Innovation state space models are stochastic processes that describe how the current state of the system depends on its **previous state** and a **random innovation term**.

- They can produce a **forecast distribution** that enables to compute **prediction intervals**.
- Their parameters are estimated with **maximum likelihood estimation**.
- The goodness of fit can be evaluated with **information criteria**.

We consider the Exponential smoothing with Trend and Seasonality (ETS) class of innovation state space models.

- The innovation term is a random error process $E_t \sim iid \mathcal{N}(0, \sigma^2)$
- Realizations can be expressed in absolute terms (**additive errors**): $\varepsilon_t = x_t - \hat{x}_{t|t-1}$
- Realizations can be expressed in relative terms (**multiplicative errors**): $\varepsilon_t = \frac{x_t - \hat{x}_{t|t-1}}{\hat{x}_{t|t-1}}$
- Choose multiplicative errors when the (residuals) variance is proportional to the level of the TS.

Simple exponential smoothing with error process

Recall the **forecast equation**

$$\hat{x}_{i+h|i} = l_i \stackrel{h=1}{\iff} \hat{x}_{i+1|i} = l_i \stackrel{i=t-1}{\iff} \hat{x}_{t|t-1} = l_{t-1}$$

With **additive errors**

$$\varepsilon_t = x_t - \hat{x}_{t|t-1} = x_t - l_{t-1}$$

Observation equation:

$$x_t = l_{t-1} + \varepsilon_t$$

Level state equation:

$$l_t = \alpha x_t + (1 - \alpha)l_{t-1} = l_{t-1} + \alpha \underbrace{(x_t - l_{t-1})}_{\varepsilon_t} = l_{t-1} + \alpha \varepsilon_t$$

with $\alpha \in [0,1]$

With **multiplicative errors**

$$\varepsilon_t = \frac{x_t - \hat{x}_{t|t-1}}{\hat{x}_{t|t-1}} = \frac{x_t - l_{t-1}}{l_{t-1}}$$

Observation equation:

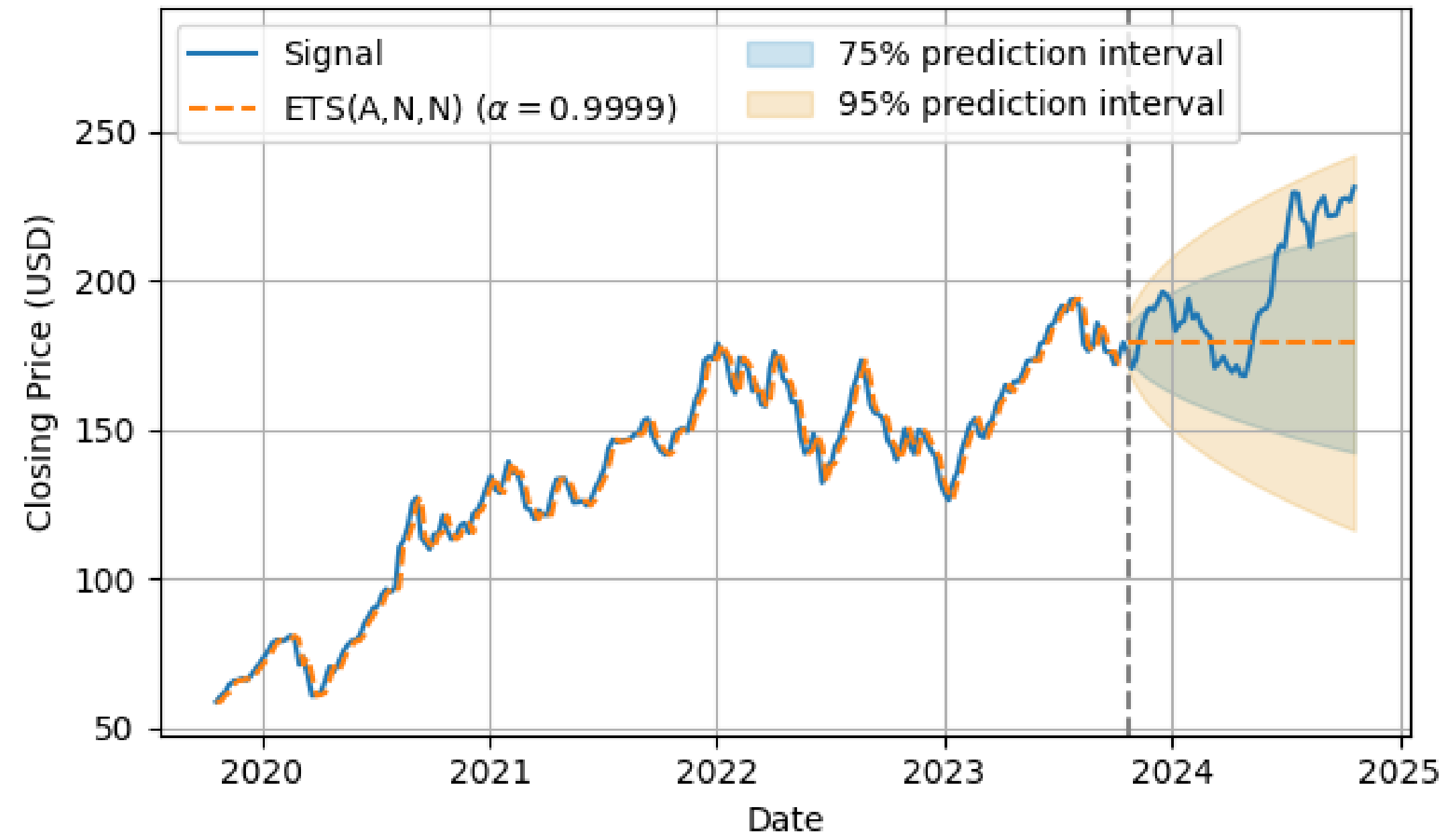
$$x_t = l_{t-1}(1 + \varepsilon_t)$$

Level state equation:

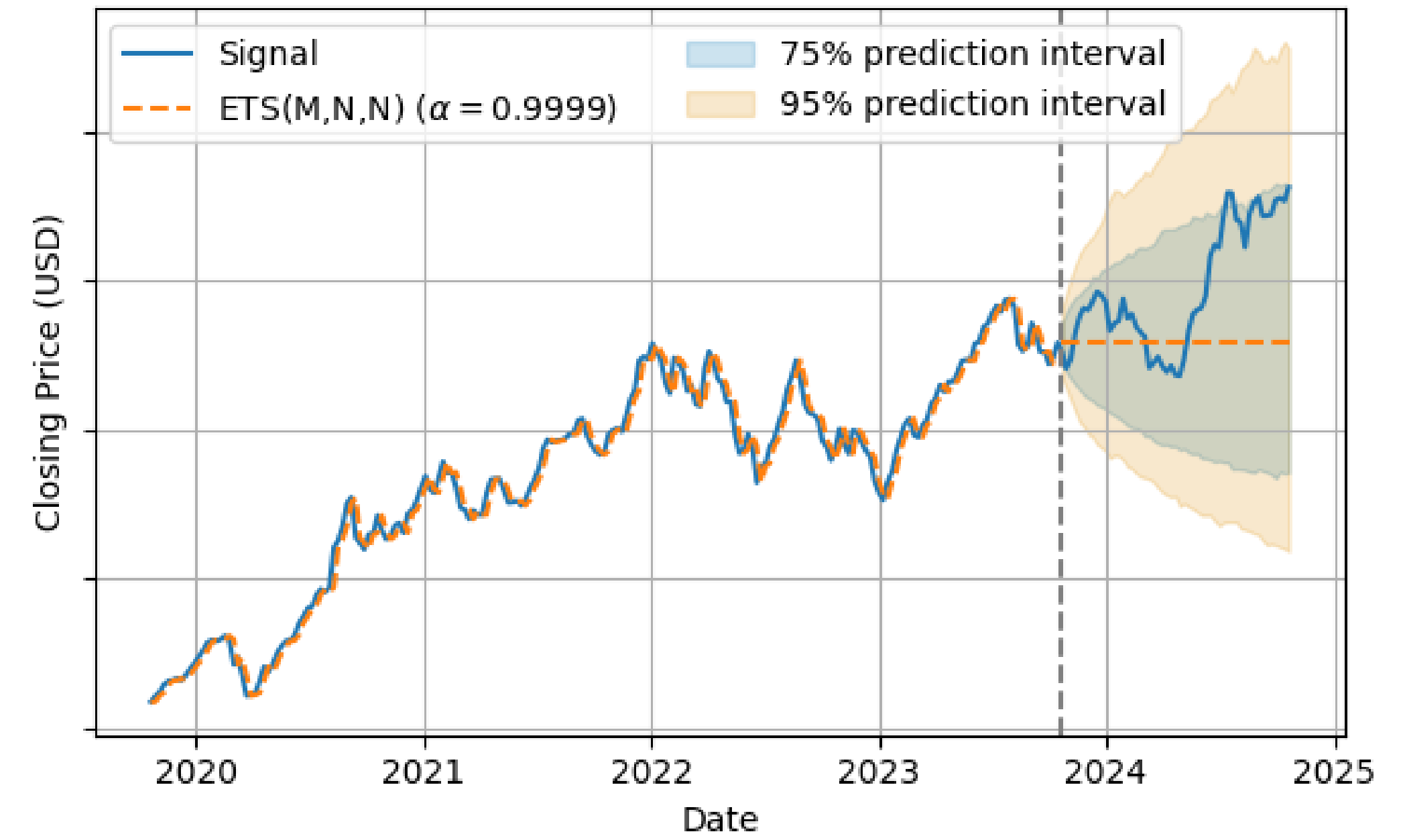
$$l_t = l_{t-1} + \alpha \underbrace{(x_t - l_{t-1})}_{\varepsilon_t l_{t-1}} = l_{t-1} + \alpha \varepsilon_t l_{t-1} = l_{t-1}(1 + \alpha \varepsilon_t)$$

with $\alpha \in [0,1]$

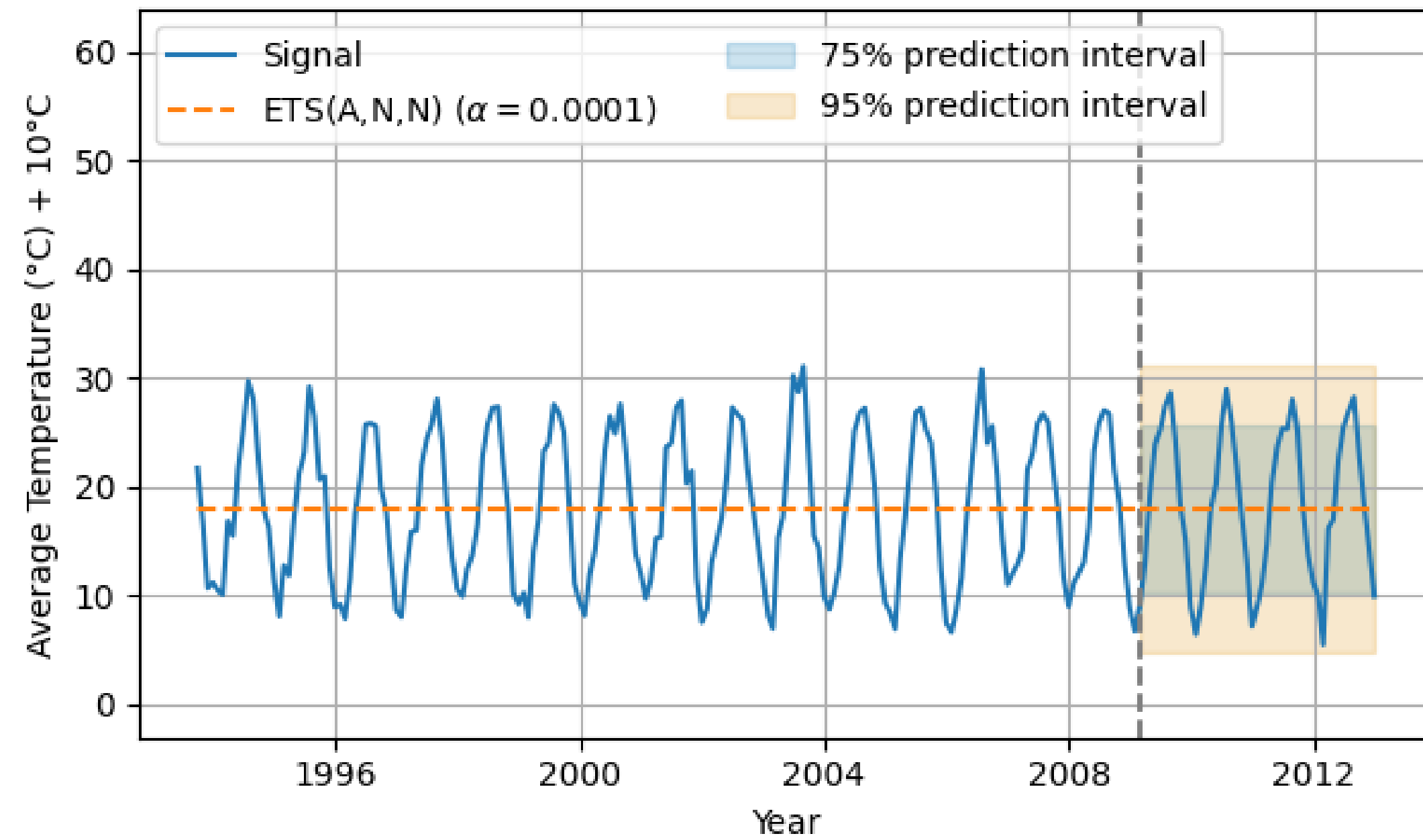
Error Type: add



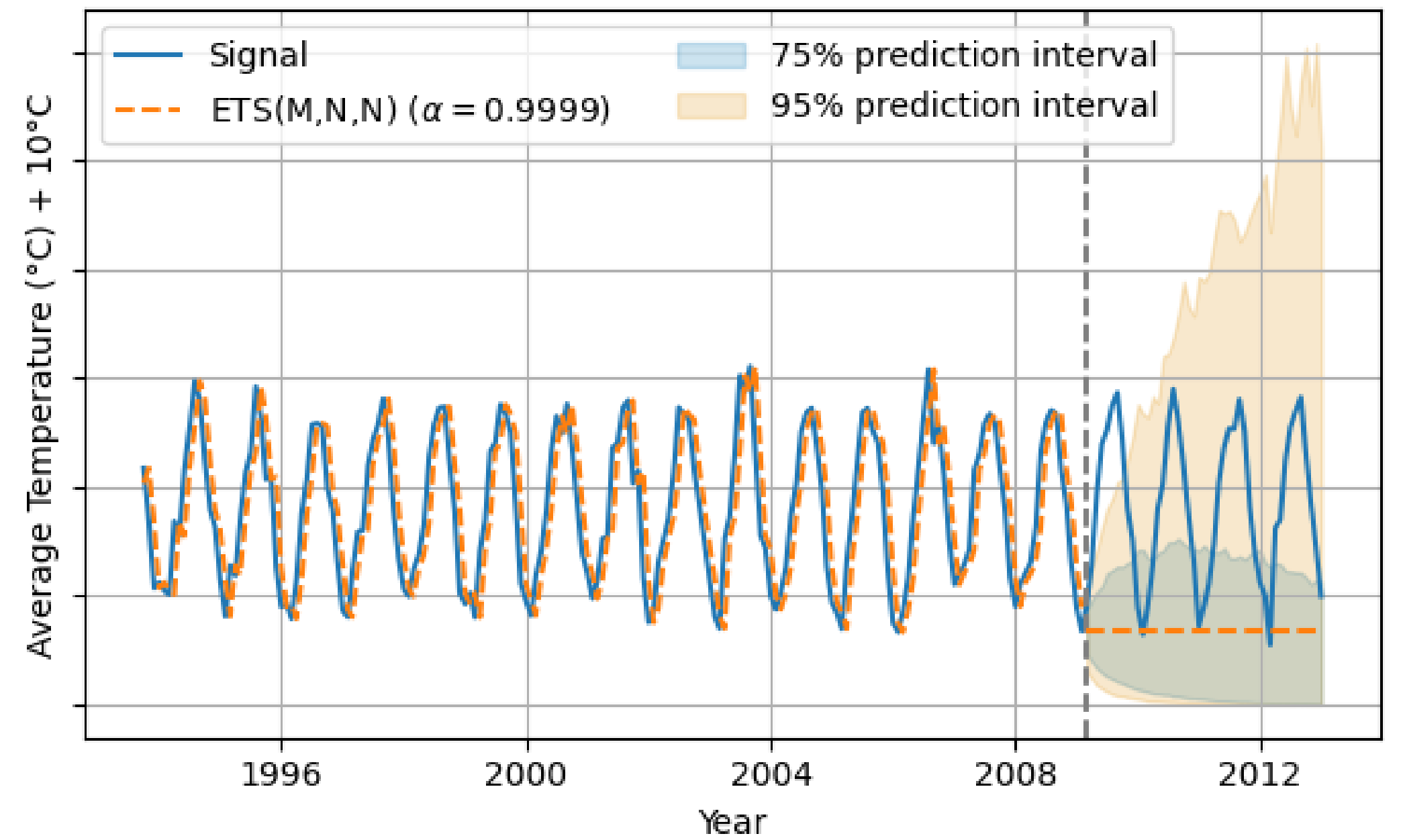
Error Type: mul



Error Type: add



Error Type: mul



Damped Holt's linear trend with error process

Exam: How to get from deterministic model to stochastic n

Recall the **forecast equation**

$$\hat{x}_{i+h|i} = l_i + b_i \sum_{j=1}^h \phi^j \stackrel{h=1}{\iff} \hat{x}_{i+1|i} = l_i + \phi b_i \stackrel{i=t-1}{\iff} \hat{x}_{t|t-1} = l_{t-1} + \phi b_{t-1}$$

With **additive errors**

$$\varepsilon_t = x_t - \hat{x}_{t|t-1} = x_t - (l_{t-1} + \phi b_{t-1})$$

Observation equation:

$$x_t = l_{t-1} + \phi b_{t-1} + \varepsilon_t$$

Level state equation:

$$l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) = l_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$$

Trend state equation:

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)\phi b_{t-1} = \phi b_{t-1} + \beta \varepsilon_t \quad \text{with } \alpha, \beta^*, \phi \in [0,1] \text{ and } \beta = \alpha \beta^*$$

With **multiplicative errors**

$$\varepsilon_t = \frac{x_t - \hat{x}_{t|t-1}}{\hat{x}_{t|t-1}} = \frac{x_t - (l_{t-1} + \phi b_{t-1})}{l_{t-1} + \phi b_{t-1}}$$

Observation equation:

$$x_t = (l_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$$

Level state equation:

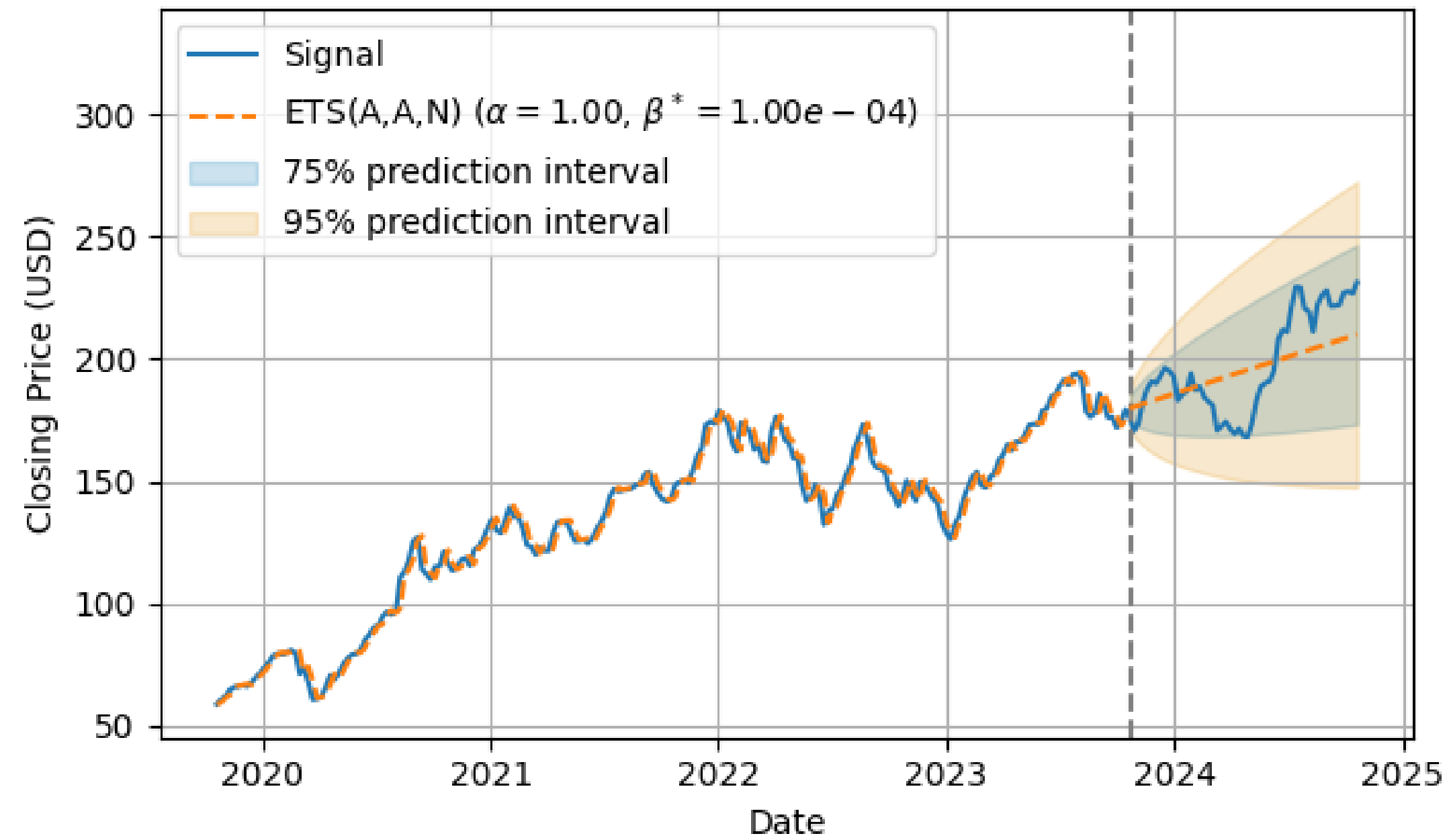
$$l_t = (l_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$$

Trend state equation:

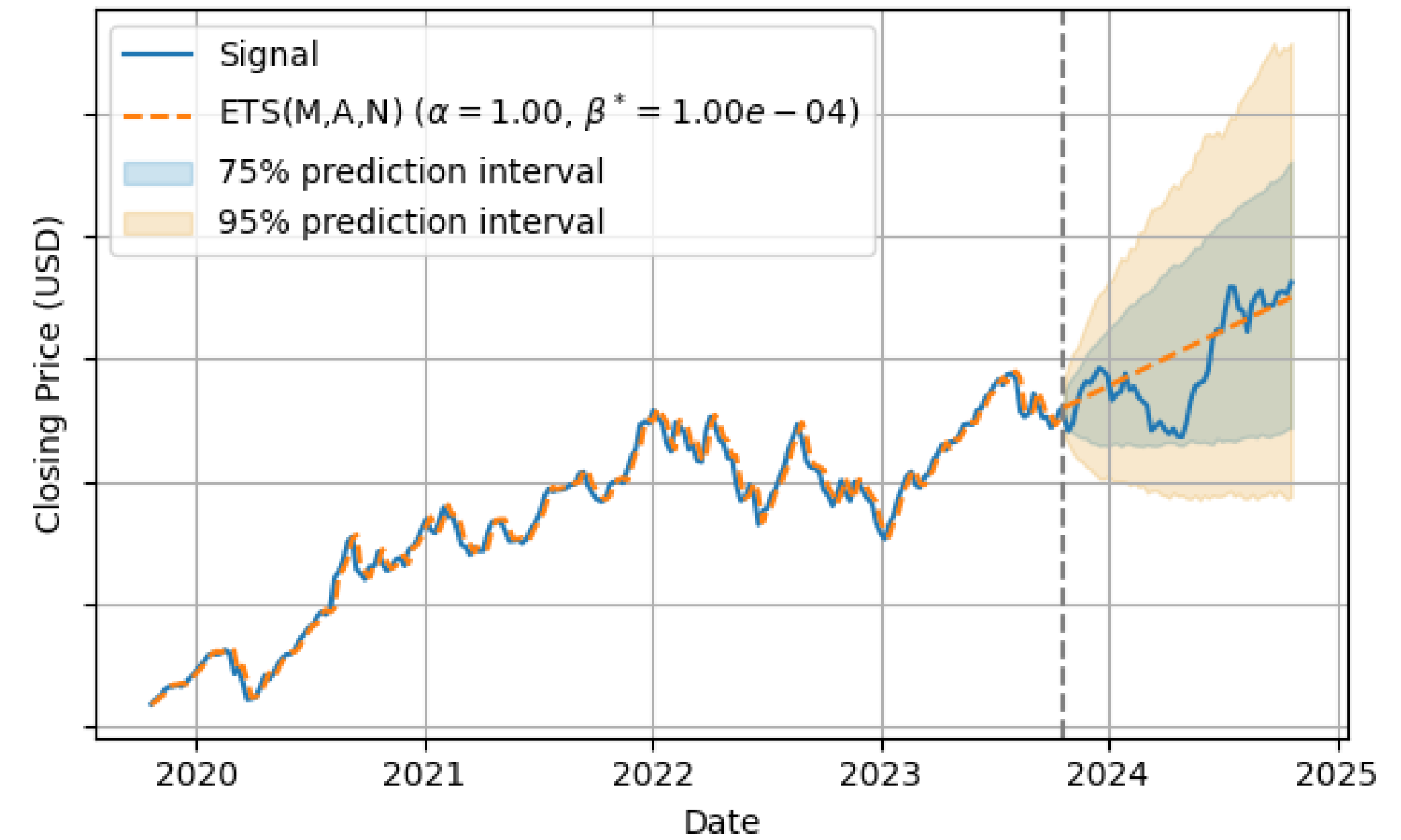
$$b_t = \phi b_{t-1} + \beta (l_{t-1} + \phi b_{t-1}) \varepsilon_t$$

with $\alpha, \beta^*, \phi \in [0,1]$ and $\beta = \alpha \beta^*$

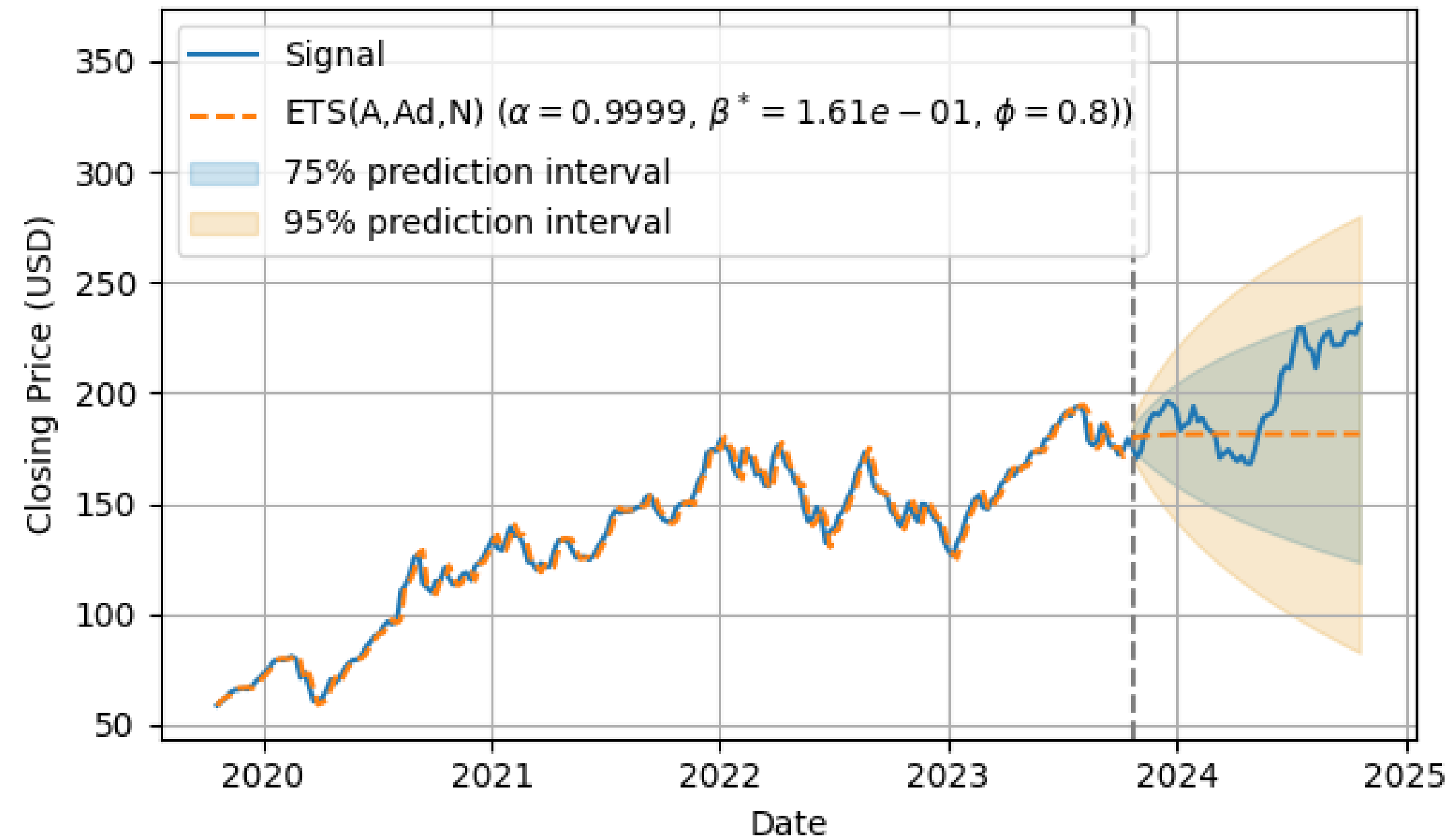
Error Type: add



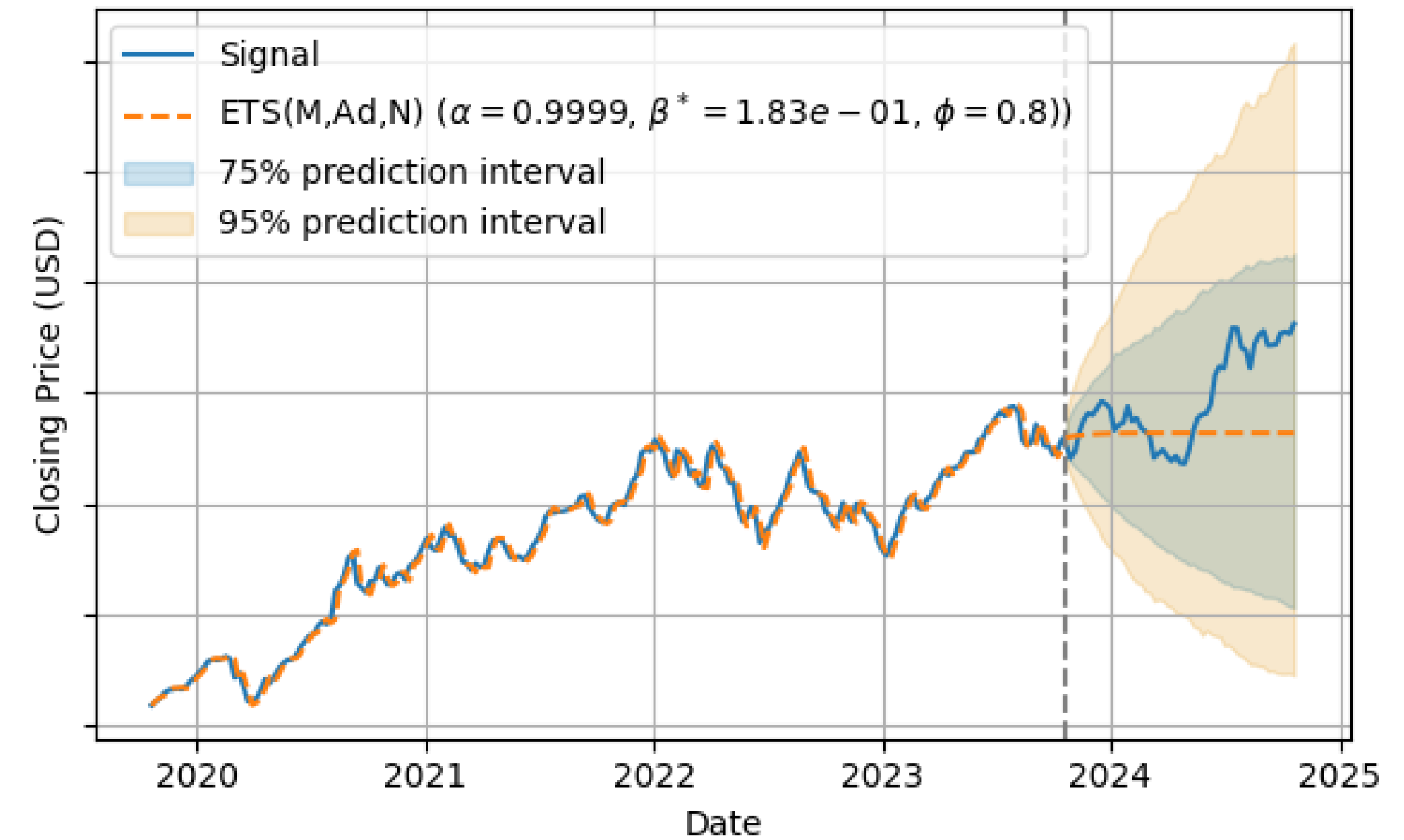
Error Type: mul



Error Type: add



Error Type: mul



Holt-Winters additive method with additive error process

Recall the **forecast equation** $\hat{x}_{i+h|i} = l_i + hb_i + s_{i+h-P[h/P]} \stackrel{h=1}{\iff} \hat{x}_{i+1|i} = l_i + b_i + s_{i+1-P} \stackrel{i=t-1}{\iff} \hat{x}_{t|t-1} = l_{t-1} + b_{t-1} + s_{t-P}$

SES with **additive errors**

$$\varepsilon_t = x_t - \hat{x}_{t|t-1} = x_t - l_{t-1} - b_{t-1} - s_{t-P}$$

Observation equation:

$$x_t = l_{t-1} + b_{t-1} + s_{t-P} + \varepsilon_t$$

Level state equation:

$$l_t = l_{t-1} + b_{t-1} + \alpha \varepsilon_t$$

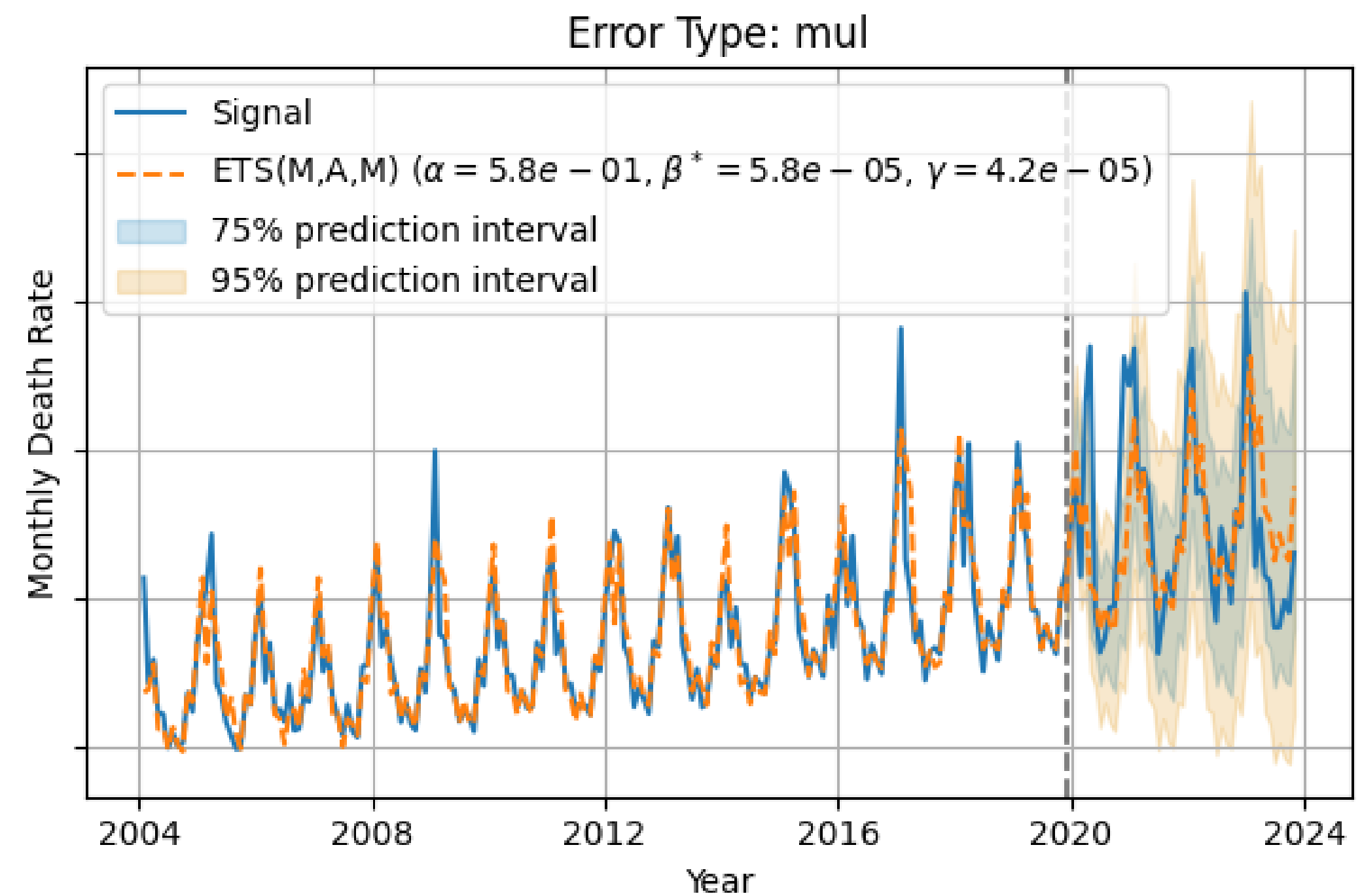
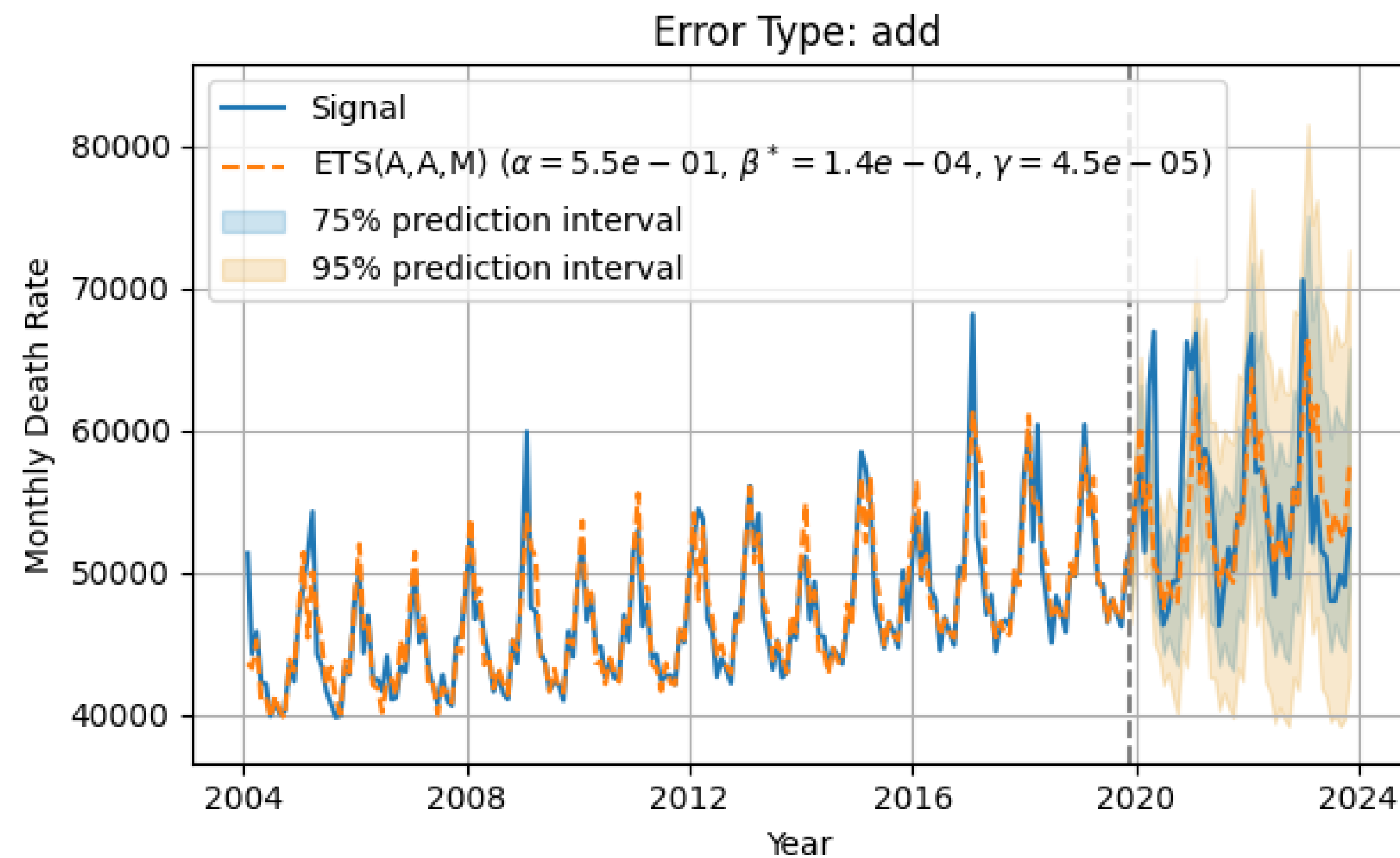
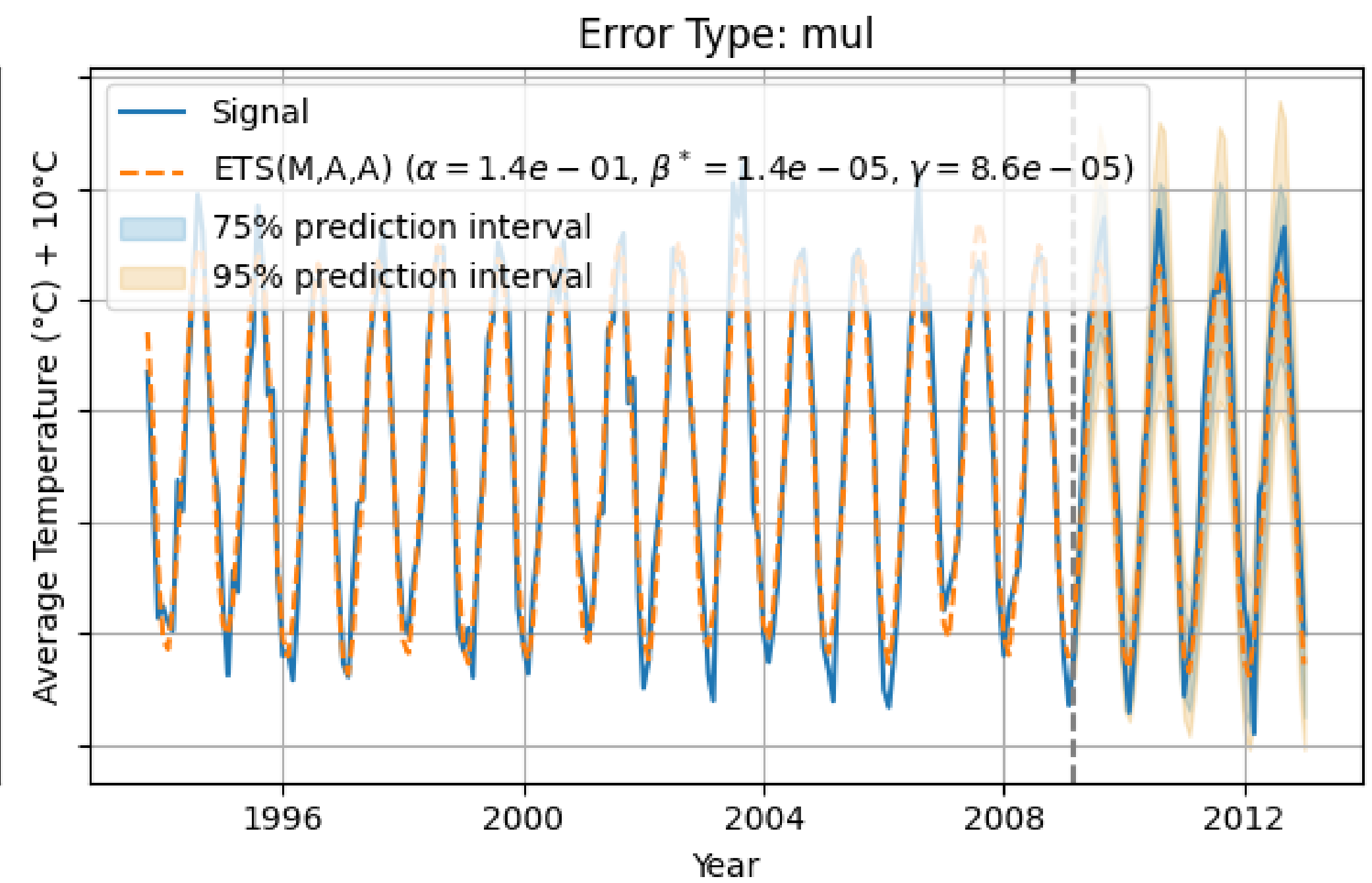
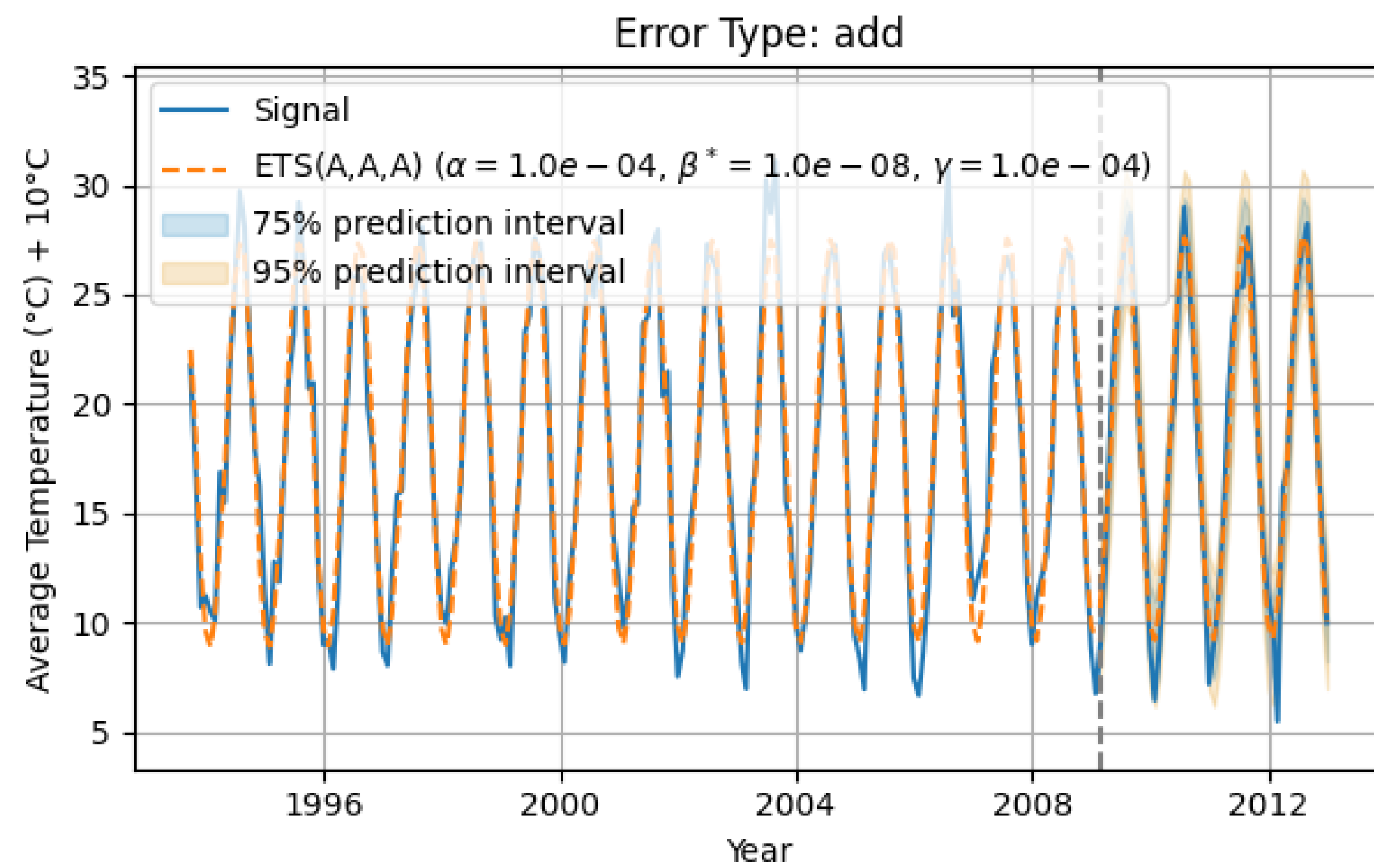
Trend state equation:

$$b_t = b_{t-1} + \beta \varepsilon_t$$

Seasonal state equation:

$$s_t = \gamma(x_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-P} = s_{t-P} + \gamma \varepsilon_t$$

with $\alpha, \beta^* \in [0,1]$, $\beta = \alpha\beta^*$ and $\gamma \in [0,1 - \alpha]$



ETS nomenclature

ETS models are labeled as ETS(\cdot, \cdot, \cdot) for respectively the error, trend, and seasonal types.

- Error is either additive (A) or multiplicative (M)
- Trend is either none (N), (A), additive damped (A_d).
- Seasonal is either (N), (A) or (M)

Multiplicative rarely ever used in practice

Multiplicative (damped) ($M_{(d)}$) trend methods are not considered since they tend to produce poor forecast.

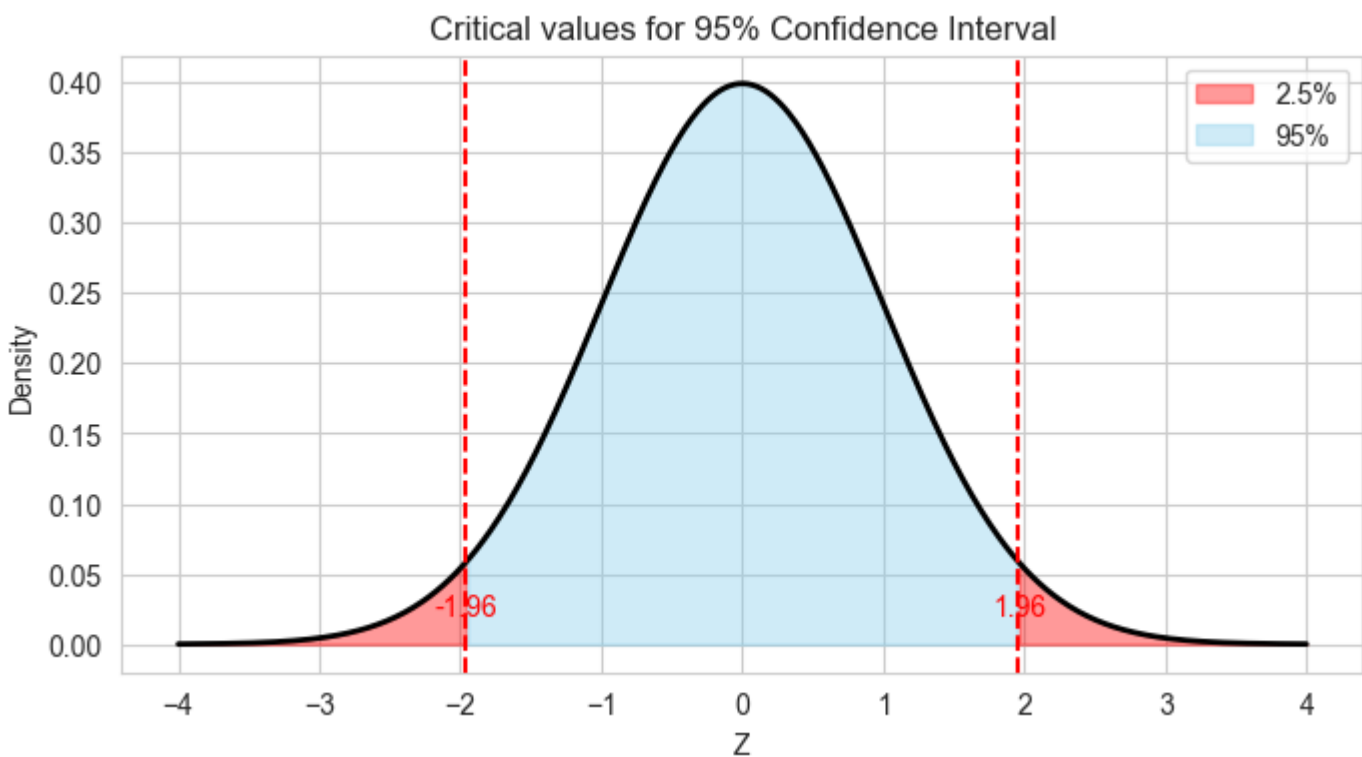
Name	Method
ETS(A,N,N)	SES with additive error
ETS(A,A,N)	Holt’s linear trend with additive error
ETS(M, A_d ,N)	Damped Holt’s linear trend with multiplicative error
ETS(M,A,A)	Additive Holt-Winters method with multiplicative error
ETS(M, A_d ,M)	Damped multiplicative Holt-Winters method with multiplicative error

Prediction confidence intervals

Confidence intervals provide a measure of **forecast uncertainty**.

When the residuals are **normally distributed**, the CI of $\hat{x}_{T+h|T}$ is $\hat{x}_{T+h|T} \pm z_{1-\alpha/2} \hat{\sigma}_{h|T}$

- $z_{1-\alpha/2}$ is the critical value of the normal distribution e.g., for 95% CI $z_{1-\alpha/2} = 1.96$
- $\hat{\sigma}_{h|T}$ is an **estimate of the standard deviation** of the h -step forecast with $\hat{\sigma}_1 = \hat{\sigma}_e$ (residuals std)



For $h > 1$, $\hat{\sigma}_{h|T}$ depends on the ETS model and how the **innovations accumulate**:

ETS(A,N,N)	ETS(A,A,N)	ETS(A,N,A)
$\hat{\sigma}_{h T}^2 = \hat{\sigma}_e^2 (1 + \alpha^2 (h - 1))$	$\hat{\sigma}_{h T}^2 = \hat{\sigma}_e^2 \left(1 + (h - 1) \left(\alpha^2 + \alpha \beta h + \frac{1}{6} \beta^2 h (2h - 1) \right) \right)$	$\hat{\sigma}_{h T}^2 = \hat{\sigma}_e^2 \left(1 + \alpha^2 (h - 1) + \gamma \left\lfloor \frac{h}{P} \right\rfloor (2\alpha + \gamma) \right)$

Use **bootstrapping** for **non-normal** residuals that are **uncorrelated** and have **constant variance**.

Connection to ARIMA (1)

Consider a zero-mean ARIMA(0,1,1) where $\theta = \alpha - 1$:

$$\begin{aligned}\Phi(B)\nabla x_t &= c + \Theta(B)w_t \\ \Leftrightarrow \nabla x_t &= w_t + (\alpha - 1)w_{t-1}\end{aligned}$$

Setting $y_t = \nabla x_t$:

$$y_t = w_t + (\alpha - 1)w_{t-1} \Leftrightarrow w_t = y_t + (1 - \alpha)w_{t-1}$$

Rewriting the model as an AR(∞):

$$\begin{aligned}w_t &= y_t + (1 - \alpha)w_{t-1} \\ &= y_t + (1 - \alpha)(y_{t-1} + (1 - \alpha)w_{t-2}) \\ &= y_t + (1 - \alpha)y_{t-1} + (1 - \alpha)^2(y_{t-2} + (1 - \alpha)w_{t-3}) \\ &= \dots \\ &= y_t + \sum_{i=1}^{\infty} (1 - \alpha)^i y_{t-i}\end{aligned}$$

Connection to ARIMA (2)

Substituting back $y_t = \nabla x_t = x_t - x_{t-1}$:

$$x_t - x_{t-1} = w_t - \sum_{i=1}^{\infty} (1 - \alpha)^i (x_{t-i} - x_{t-i-1})$$

Expanding the sum, choosing two different indexes for clarity:

$$\begin{aligned} x_t &= x_{t-1} + w_t - \sum_{i=1}^{\infty} (1 - \alpha)^i x_{t-i} + \sum_{j=1}^{\infty} (1 - \alpha)^j x_{t-j-1} \\ &= x_{t-1} + w_t - \sum_{i=1}^{\infty} (1 - \alpha)^i x_{t-i} + \sum_{k=2}^{\infty} (1 - \alpha)^{k-1} x_{t-(k-1)-1} & (j = k - 1) \\ &= w_t - \sum_{i=1}^{\infty} (1 - \alpha)^i x_{t-i} + \sum_{k=1}^{\infty} (1 - \alpha)^{k-1} x_{t-k} \\ &= w_t + \sum_{i=1}^{\infty} (1 - \alpha)^{i-1} x_{t-i} (1 - (1 - \alpha)) = w_t + \sum_{i=1}^{\infty} \alpha (1 - \alpha)^{i-1} x_{t-i} & (k = i) \end{aligned}$$

Connection to ARIMA (3)

Consider the 1-step ahead prediction:

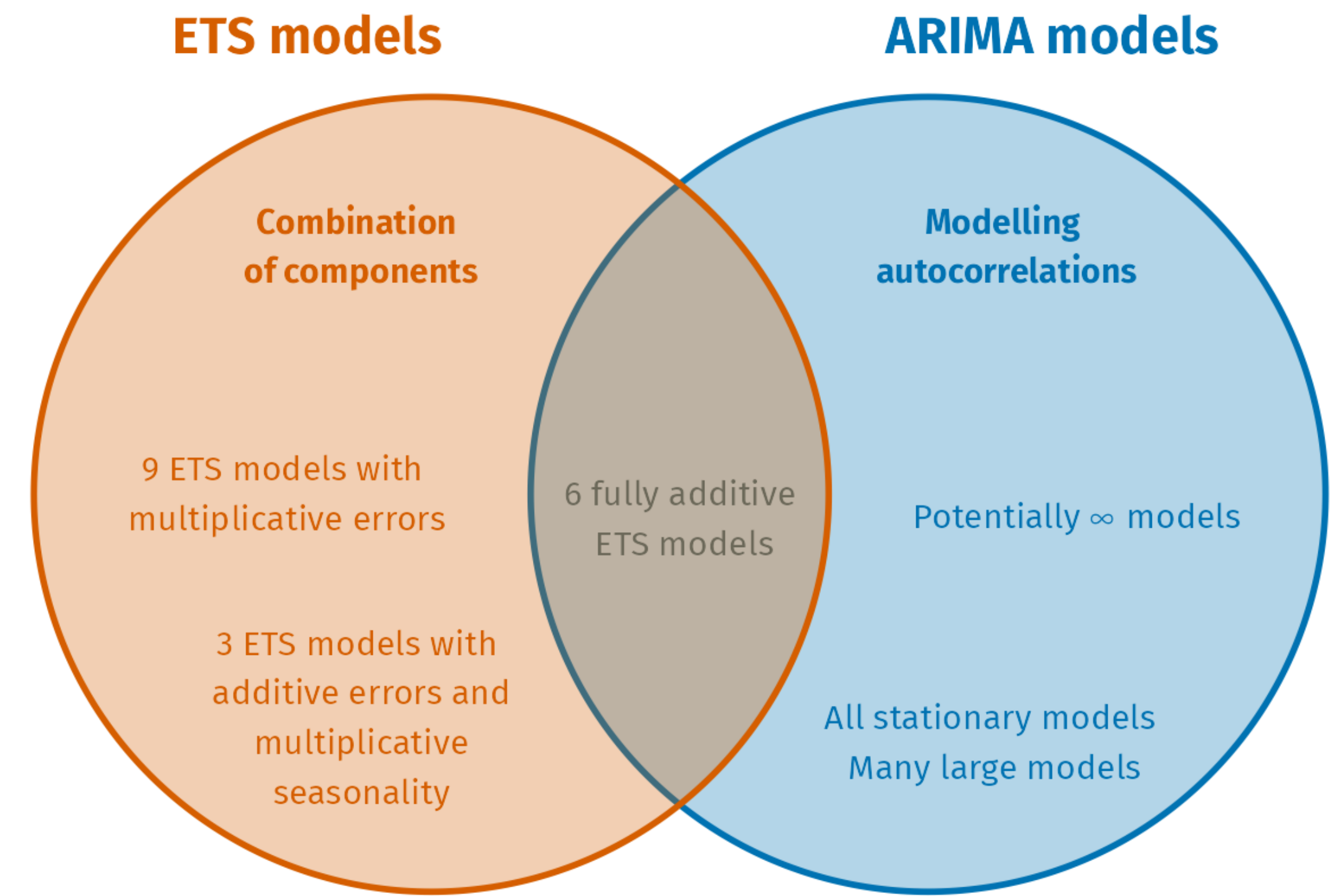
$$\begin{aligned}\hat{x}_{t+1|t} &= \underbrace{w_{t+1}}_0 + \sum_{i=1}^{\infty} \alpha(1-\alpha)^{i-1} x_{t+1-i} \\ &= \alpha x_t + \sum_{i=2}^{\infty} \alpha(1-\alpha)^{i-1} x_{t+1-i} \\ &= \alpha x_t + \sum_{k=1}^{\infty} \alpha(1-\alpha)^k x_{t-k} && (k = i - 1) \\ &= \alpha x_t + (1-\alpha) \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} x_{t-k} \\ &= \alpha x_t + (1-\alpha) \hat{x}_{t|t-1}\end{aligned}$$

$\hat{x}_{t+1|t}$ is the avg_w of the current observation x_t and the previous prediction $\hat{x}_{t|t-1}$.
Thus, the zero-mean ARIMA(0,1,1) is **equivalent** to simple exponential smoothing.

Connection to ARIMA (4)

ETS models are non-stationary while some ARIMA models are stationary.

- ETS models with seasonality or non-damped trend have two unit-roots i.e., require two levels of differencing.
- Other ETS models have one unit root i.e., require one level of differencing.



Fully additive ETS	ARIMA equivalent
ETS(A,N,N)	ARIMA(0,1,1)
ETS(A,A,N)	ARIMA(0,2,2)
ETS(A,A _d ,N)	ARIMA(1,1,2)
ETS(A,N,A)	ARIMA(0,1,P)(0,1,0) _P
ETS(A,A,A)	ARIMA(0,1,P+1)(0,1,0) _P
ETS(A,A _d ,A)	ARIMA(1,0,P+1)(0,1,0) _P

Modelling different components

Modelling autocorrelation between

Exercise

Review statsmodels [notebook](#) on exponential smoothing.

Simulate and analyze different synthetic signals

- Compare variations of ETS error, trend, seasonality
- Confirm equivalence with ARIMA models when applicable.
- Estimate the model parameters from the generated data.

Model real-world time series

- Compare ETS models with information criteria and residual analysis
- Compare with ARIMA performance using cross-validation
- Compare prediction intervals
- Compare ETS components with decomposition methods