ILLINOIS
Gies College of Business

Introduction to Forecasting

Average Method

Dr. Sung Won KIm

Average Method

Forecasts of future value is the mean of the past data.

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \cdots + y_T)/T$$

ILLINOIS
Gies College of Business

Introduction to Forecasting

Naïve Method

Dr. Sung Won Kim

Naïve Method

The naïve method uses the most recent observation as the forecast.

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

ILLINOIS
Gies College of Business

Introduction to Forecasting

Linear Regression

Dr. Sung Won Kim

Linear Model – Introduction

Regression analysis – use a model to examine the relationship between a variable Y and other variables $X_1, \dots, X_p$.

Y is called the response variable.

$X_1, \dots, X_p$ are called the predictors.

Linear Regression

For linear regression, we use a linear equation to build the relationship. We use Y_t to denote the value of the response at time t, and we have:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \varepsilon_t$$
$$= \beta_0 + \sum_{i=1}^{k} \beta_i X_{it} + \varepsilon_t$$

The βs are called coefficients, and β_0 is called the **intercept**. εs are called **error terms**.

Assumptions

For the model we build, we have the following assumptions:

Response and predictors have **linear** relationship.

Independent variables are multivariate normal.

Independent variables have no or little multicollinearity.

No auto correlation. The error terms are independent.

$$\left(cov\left(\varepsilon_i, \varepsilon_j\right) = 0, i \neq j\right)$$

The error terms are i.i.d. normal with a zero mean the same variance.

$$\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$$

Simple Linear Regression and Estimation

The simple linear regression model only contains one predictor:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \qquad t = 1, 2, \ldots, n$$

Estimate parameters β_0, β_1 by minimizing errors:

$$\min_{\beta_0, \beta_1} \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} (y_t - \beta_0 - \beta_1 x_t)^2 \ ,$$

which is called the **least squares** criterion.

Simple linear regression and estimation

The estimates $\hat{\beta}_0, \hat{\beta}_1$ are given by the formula:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

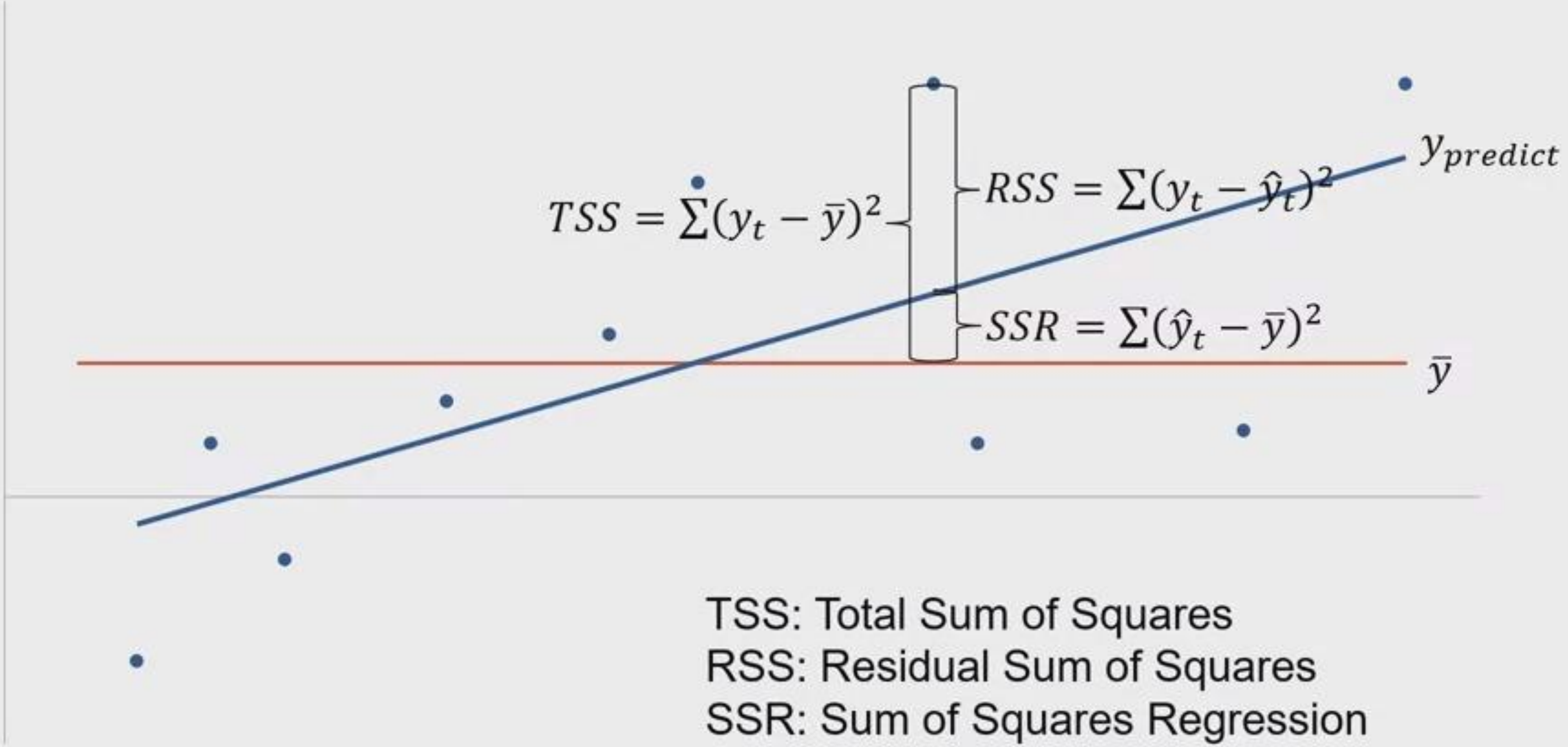
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Then we have the estimates for response and error:

$$\hat{y}_{\mathrm{t}} = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

$$\hat{\varepsilon}_t = y_t - \hat{y}_t$$

Illustration



TSS: Total Sum of Squares
RSS: Residual Sum of Squares
SSR: Sum of Squares Regression

Sum of Squares

Residual sum of squares(RSS): the sum of the squared least-squares residuals

$$RSS = \sum_{t=1}^{n} \hat{\varepsilon}_t^{\,2}$$

Total sum of squares(TSS):

$$TSS = \sum_{t=1}^{n} (y_t - \bar{y})^2$$

Sum of squares regression(SSR):

$$SSR = \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2$$

Degrees of Freedom and R-Squared

Degrees of freedom: the number of observations minus the number of regression predictors in model

R^2: coefficient of determination, which can also be viewed as the percentage of information explained by the model

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

(Note: $0 \leq R^2 \leq 1$)

ILLINOIS
Gies College of Business

Moving Average

Dr. Sung Won Kim

Simple Moving Average

The *k*-period **moving average** is a series of average values calculated using consecutive time periods. It is a smoothing technique to get an overall idea of the trend.

Simple moving average: $A_t = \frac{1}{k}\sum_{t-k+1}^{t} x_t$

Where x_t are the observations, A_t is the moving average at time, *t*. The moving average uses *n* consecutive time periods.

Modified from Forecasting: Principles and Practice: https://otexts.com/fpp2/moving-averages.html

Calculating a Simple Moving Average

time	sales	3 lags	5 lags			
10	27					
9	15					
8	18					
7	12					
6	18					
5	17					
4	23					
3	22					
2	19					

Moving Averages

The moving average is not good for data with no trend or seasonality.

ILLINOIS
Gies College of Business

Introduction to Exponential Smoothing

Dr. Sung Won Kim

Simple Moving Average of Order k

$$A_t = \frac{1}{k} \sum_{t-k+1}^{t} x_t$$

$$A_t = \frac{(x_t + \ x_{t-1} + \ x_{t-2} + \cdots + x_{t-k+1}\ )}{k}$$

If there is no trend or seasonality, then the moving average is a good forecast.

Modified from Forecasting: Principles and Practice: https://otexts.com/fpp2/moving-averages.html

k: Order of the Moving Average

If k is **small**, the **more** weight we place on recent events.

A **small** k is useful when there are sudden changes in the data.

If k is **large**, the **less** weight we place on recent events.

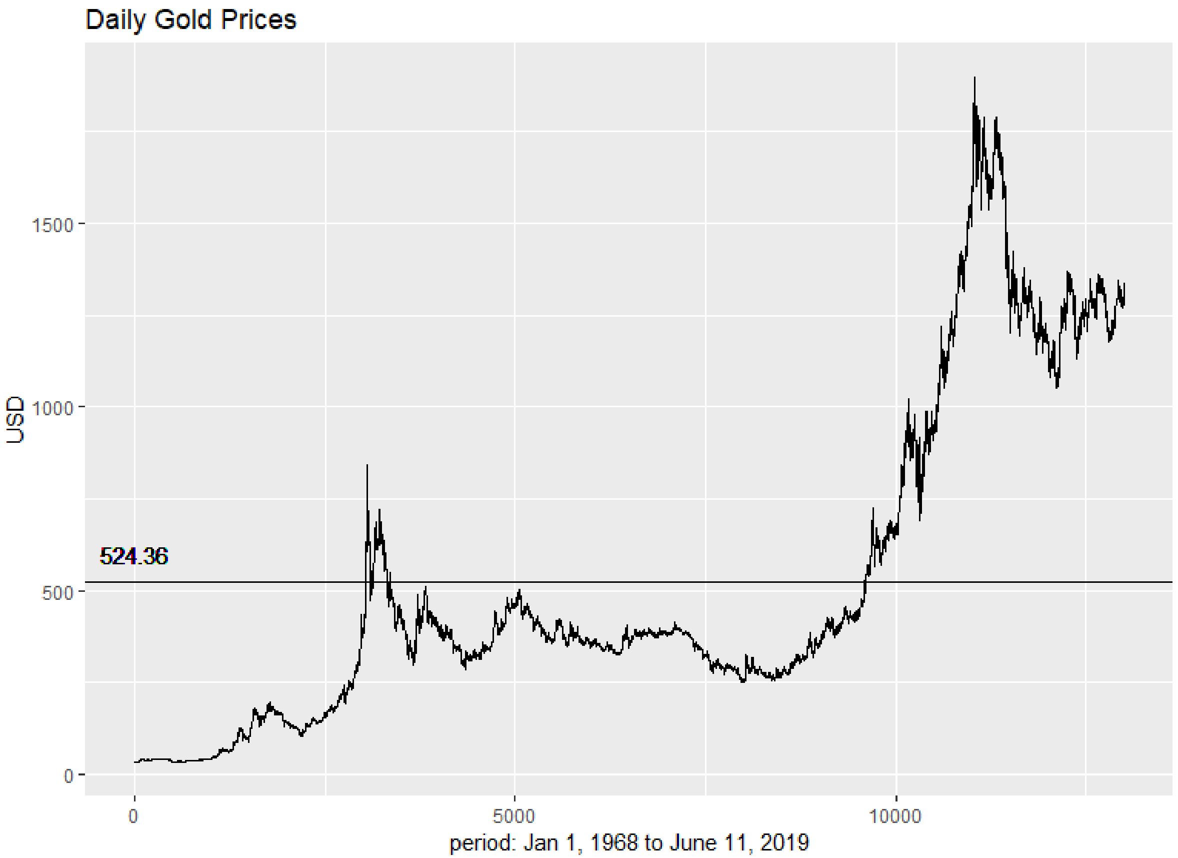
A **large** k is useful when there are infrequent changes over large periods in the data.

k: Order of the Moving Average

If $k = 1$, then we use the current period to forecast the next period.

If $k = N$ (i.e., we use all the data points), then this is ok if there is no trend or seasonality.

Consider When to Use Large and Small k



Exponential Smoothing Methods

Exponential smoothing methods give more weight to more recent events.

Good for data with trends or seasonal cycles

https://en.wikipedia.org/wiki/Exponential_smoothing

Three Smoothing Methods

Simple exponential smoothing – puts more weight on recent past to smooth data

Holt’s exponential smoothing – allows for data with a trend

Holt-Winters’ method – allows for data with a trend **and** seasonal cycles

ILLINOIS
Gies College of Business

Simple Exponential Smoothing

Dr. Sung Won KIm

Simple Exponential Smoothing

Used past data points to forecast the future like moving averages

Exponential gives more weight to more recent values.

Past values are smoothed like moving averages.

Simple Exponential Smoothing

$$S_t = \alpha x_t + (1 - \alpha) s_{t-1}$$

$$0 < \alpha < 1$$

α is the smoothing factor.

Simple Exponential Smoothing

$$s_0 = x_0$$
$$s_t = \alpha x_t + (1-\alpha)s_{t-1}; t > 0$$

α = smoothing factor (0<α<1)

s_t = period t's forecast value

x_t = actual value in time t

s_{t-1}= forecast value for $t-1$

Why Is It Called “Exponential”?

$$s_t = \alpha x_t + (1-\alpha)s_{t-1}$$

Why Is It Called “Exponential”?

$$s_t = \alpha x_t + (1-\alpha)s_{t-1}$$

$$s_{t-1} = \alpha x_{t-1} + (1-\alpha)s_{t-2}$$

So

$$\begin{aligned} s_t &= \alpha x_t + (1-\alpha)[\alpha x_{t-1} + (1-\alpha)s_{t-2}] \\ &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2 s_{t-2} \end{aligned}$$

More Generally…

$$s_t = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} \\ + \alpha(1-\alpha)^3 x_{t-3} + \cdots + (1-\alpha)^{t-1} s_1$$

We could go back to the beginning of the data.

The weights get smaller over time.

The weights sum to 1.

How Do the Weights Behave Over Time?

Time	$\alpha = 0.2$ Calculation	Value of Weight
t		0.2
t-1	0.2 x 0.8	.16
t-2	0.2 x 0.8^2	0.128
t-3	0.2 x 0.8^3	0.102
…		
Total		1.00

Now we can see s_t is a weighted average of the observations of the prior periods.

Adapted from Wilson & Keating, 1998

How Do We Pick α?

$0 < \alpha < 1$

Try different values of alpha (.1, .2, …, .9) and compute the RMSE.

Choose the α with the lowest RMSE.

ILLINOIS
Gies College of Business

Simple Exponential Smoothing

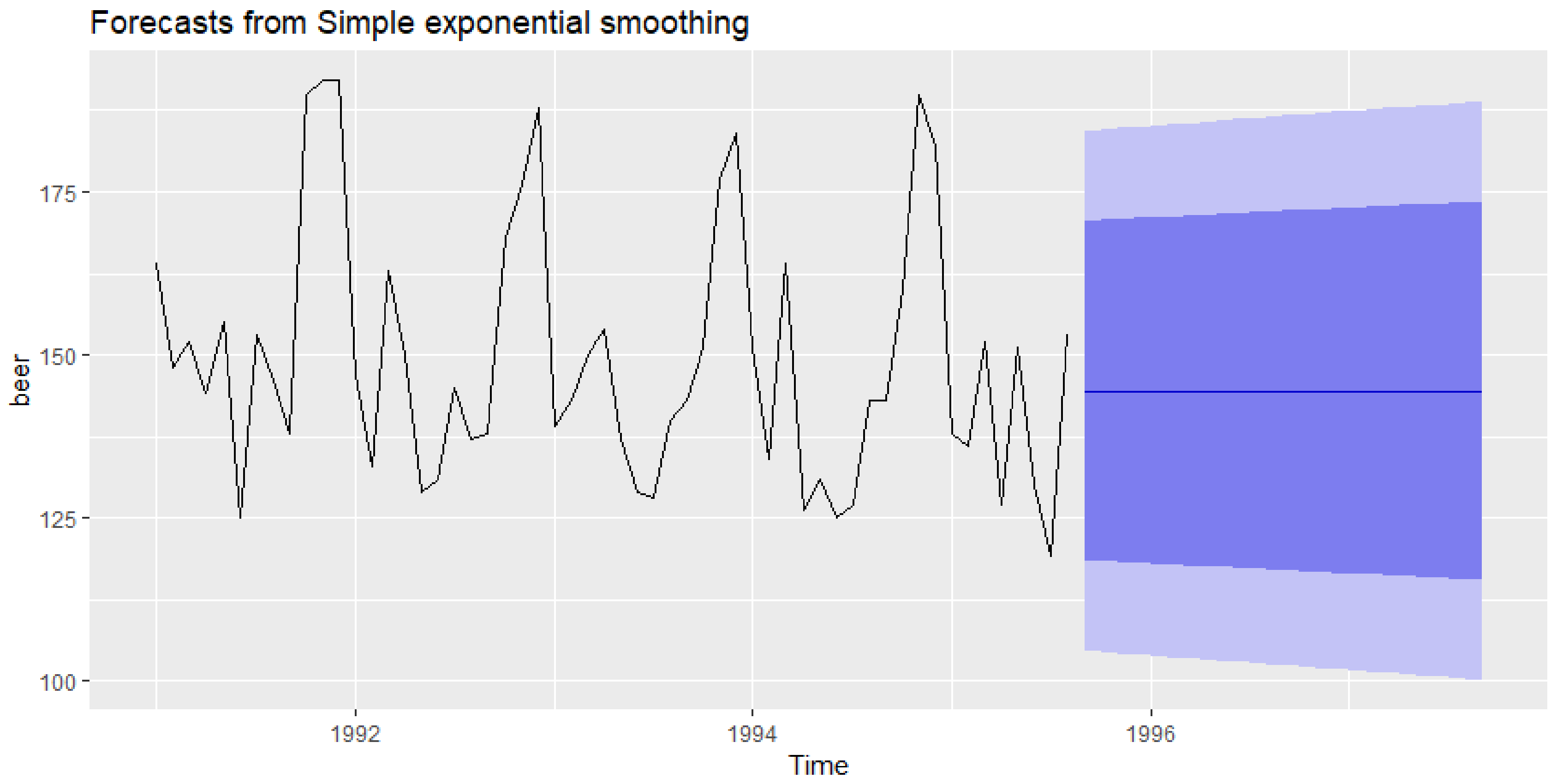
R Example

Dr. Sung Won KIm

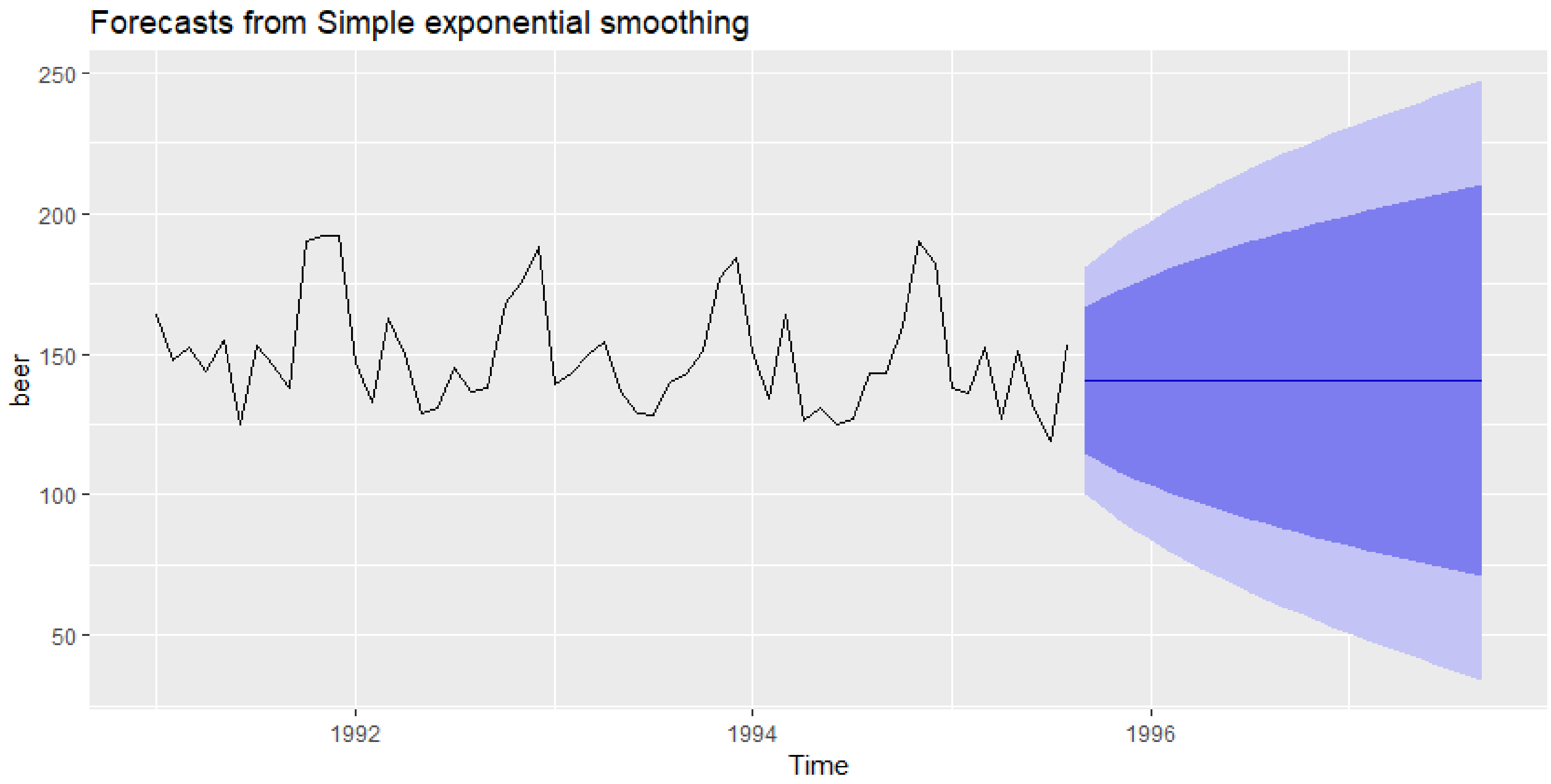
R Code Example

See example code:
Exponential smoothing.r

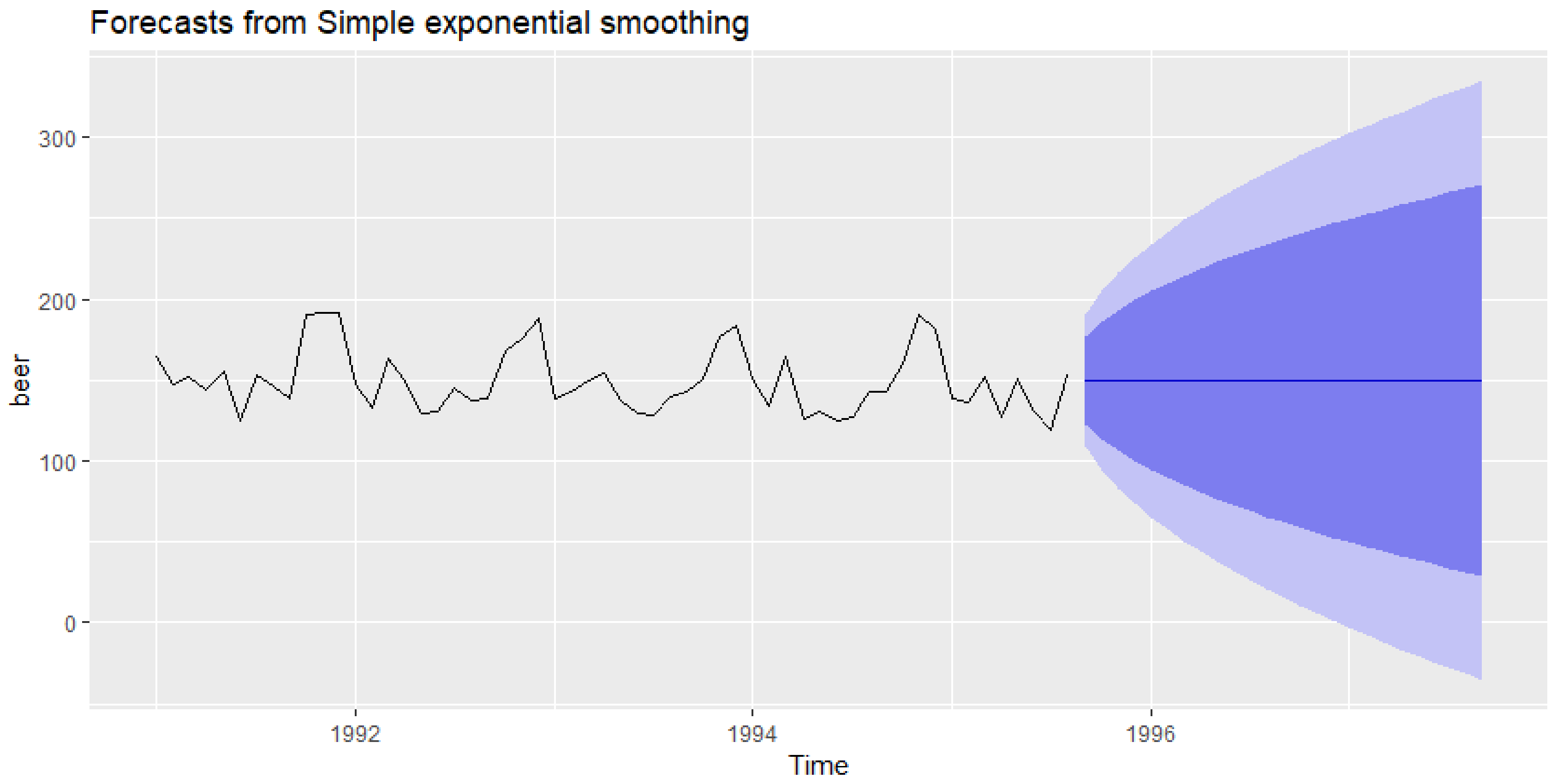
Alpha 0.1; Order 25



Alpha 0.5; Order 25



Alpha 0.9; Order 25



RMSE – Beer Example

alpha	RMSE
0.1	19.96311
0.5	20.15516
0.9	20.49995

Number of lags: 25

Recap

Exponential smoothing is best for short-term forecasts without trend or seasonality.

You need to pick alpha, the smoothing constant.

Recap

A bigger alpha means more weight is given to recent past data points.

One method to select alpha is to minimize the RMSE.

ILLINOIS
Gies College of Business

Holt’s Exponential Smoothing

Dr. Sung Won Kim

Limitations of Exponential Smoothing

The simple exponential smoothing method does not perform well with data that has a trend or seasonality

Extensions to Simple Exponential Smoothing

Holt’s exponential smoothing is used for data with trends

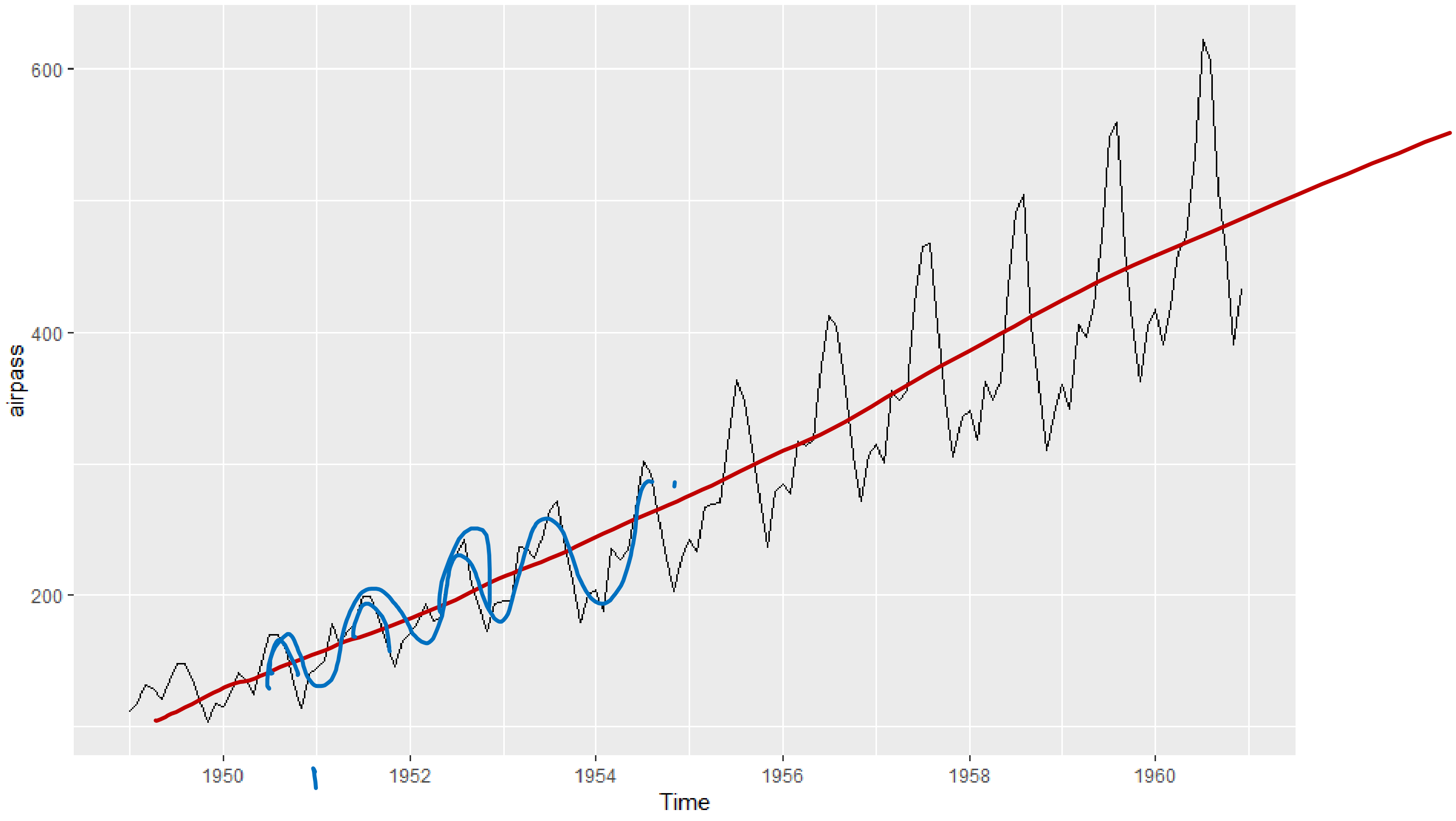
Winter’s exponential smoothing adds seasonality into the model

Monthly Airline Passengers (in thousands): 1949–1960

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1949	112	118	132	129	121	135	148	148	136
1950	115	126	141	135	125	149	170	170	158
1951	145	150	178	163	172	178	199	199	184
1952	171	180	193	181	183	218	230	242	209
1953	196	196	236	235	229	243	264	272	237
1954	204	188	235	227	234	264	302	293	259
1955	242	233	267	269	270	315	364	347	312
1956	284	277	317	313	318	374	413	405	355
1957	315	301	356	348	355	422	465	467	404
1958	340	318	362	348	363	435	491	505	404
1959	360	342	406	396	420	472	548	559	463
1960	417	391	419	461	472	535	622	606	508

R Dataset: AirPassengers

Monthly Airline Passengers: 1949–1960



Simple Exponential Smoothing

$$\hat{y}_{T+1} = \alpha y_T + (1-\alpha)\hat{y}_{T-1}$$

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_{t-1}$$

60

α = smoothing factor $(0<\alpha<1)$

$t = 1, \ldots, T$

$\hat{y}_{T+1}$ = the forecast for T+1

SES – Component Form

Forecast equation:

$$\hat{y}_{t+h} = l_t$$

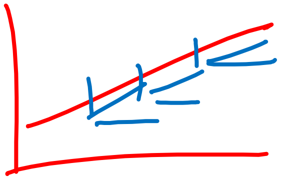
Level equation:

$$l_t = \alpha y_t + (1 - \alpha) l_{t-1}$$

l_t is the level (smoothed value).

$h = 1$ gives the fitted values.

Holt's Linear Trend



Forecast equation: $\hat{y}_{t+h} = l_t + hb_t$

Level equation: $l_t = \alpha y_t + (1-\alpha)l_{t-1}$

Trend equation: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$

l_t is the level (smoothed value).

h is the number of steps ahead.

b_t is the weighted average of the trend.

https://otexts.com/fpp2/holt.html

ILLINOIS
Gies College of Business

Holt-Winter's Forecasting Model

Dr. Sung Won Kim

Holt-Winters Additive Method

Forecast: $\hat{y}_{t+h} = l_t + hb_t + s_{t+h-m(k+1)}$

Level: $l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$

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

Seasonality: $s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$

l_t is the level (smoothed value).

h is the number of steps ahead.

b_t is the weighted average of the trend.

s_t is the seasonality estimate.

https://otexts.com/fpp2/holt.html

Holt-Winters Multiplicative Method

The Holt-Winter’s additive method is useful when the seasonal variation is constant.

The multiplicative method is useful when the seasonal variation changes in proportion to the level of the time series.

Holt-Winters Multiplicative Method

Forecast: $\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)}$

Level: $l_t = \alpha\left(\frac{y_t}{s_{t-m}}\right) + (1-\alpha)(l_{t-1} + b_{t-1})$

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

Seasonality: $s_t = \gamma\frac{y_t}{(l_{t-1}+b_{t-1})} + (1-\gamma)s_{t-m}$

l_t is the level (smoothed value).

h is the number of steps ahead.

b_t is the weighted average of the trend.

s_t is the seasonality estimate.

Holt-Winters Multiplicative Method

In the additive model, seasonality is calculated in absolute terms. The level equation accounts for seasonality through subtraction.

In the multiplicative model, seasonality is calculated in relative terms – as a percentage – and enters the level equation through division.

Holt-Winters Multiplicative Method

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)}$$

$$l_t = \alpha\left(\frac{y_t}{s_{t-m}}\right) + (1-\alpha)(l_{t-1} + b_{t-1})$$

In the level equation, y_t is divided by s_{t-m} to remove any seasonal effects in the time series.

s_{t-m} > 1 when values in (t−m) are greater than average.

Dividing by s_{t-m} reduces the value by a percentage equivalent to the percentage that the seasonal trend was over the average.

ILLINOIS

Gies College of Business

Autoregression

Introduction

Dr. Sung Won Kim

What Is Regression?

Simplest type of predictive analysis

Expressed as,

$$y = mx + c$$

where,

y = Output/Dependent variable we want to predict

x = Input/Independent variable we use to predict y

m = Regression coefficient

c = Constant

Regression vs. Autoregression

For regression we need an input variable to predict the output variable. Example – How study time (input variable) will impact marks obtained in examination (output variable)

Autoregression is a process to find relationship with itself. Example – How yesterday’s inflation (input variable) will impact today’s inflation (output variable)

What Is Autoregression?

Representation of time-varying processes

Used to explain linear dependence of any variable’s future value to its previous time-step value

In finance, autoregression is a crucial tool to analyze time-series data.

In Mathematical Terms

AR(1) is an autoregressive model with order or lag 1 defined as,

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t$$

where, φ_1 is the model parameter with ε_t as white noise.

Similarly, an AR(2) model can be written as,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

https://en.wikipedia.org/wiki/Autoregressive_model

In Mathematical Terms

In general terms, AR(p) autoregressive model with order/lag p can be written as,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

where, φ_1,……..,φ_p are model parameters ε_t as white noise.

https://en.wikipedia.org/wiki/Autoregressive_model

White Noise

$$\varepsilon_t \sim N(0, \sigma_\varepsilon), \text{for all } \sigma_\varepsilon > 0$$

$$cov(\varepsilon_u, \varepsilon_t) = 0, u \neq t$$

ε's are independent of each other.

R CODE

WHITE NOISE