Economics 144
Economic Forecasting

Lecture 3
Modeling and Forecasting Trend
(Part II)

Dr. Randall R. Rojas

Today's Class

- Forecasting Challenges
- Forecasting Environments
- Model Selection
  - MSE
  - AIC
  - SIC
- Trend Fitting via Periodic Functions
- Trend Fitting via Holt-Winters Filtering
- R Example

Forecasting Environments

- The data sample is divided into two parts: usually 2/3 are used for estimation and 1/3 for prediction.
- Def: Estimation Sample

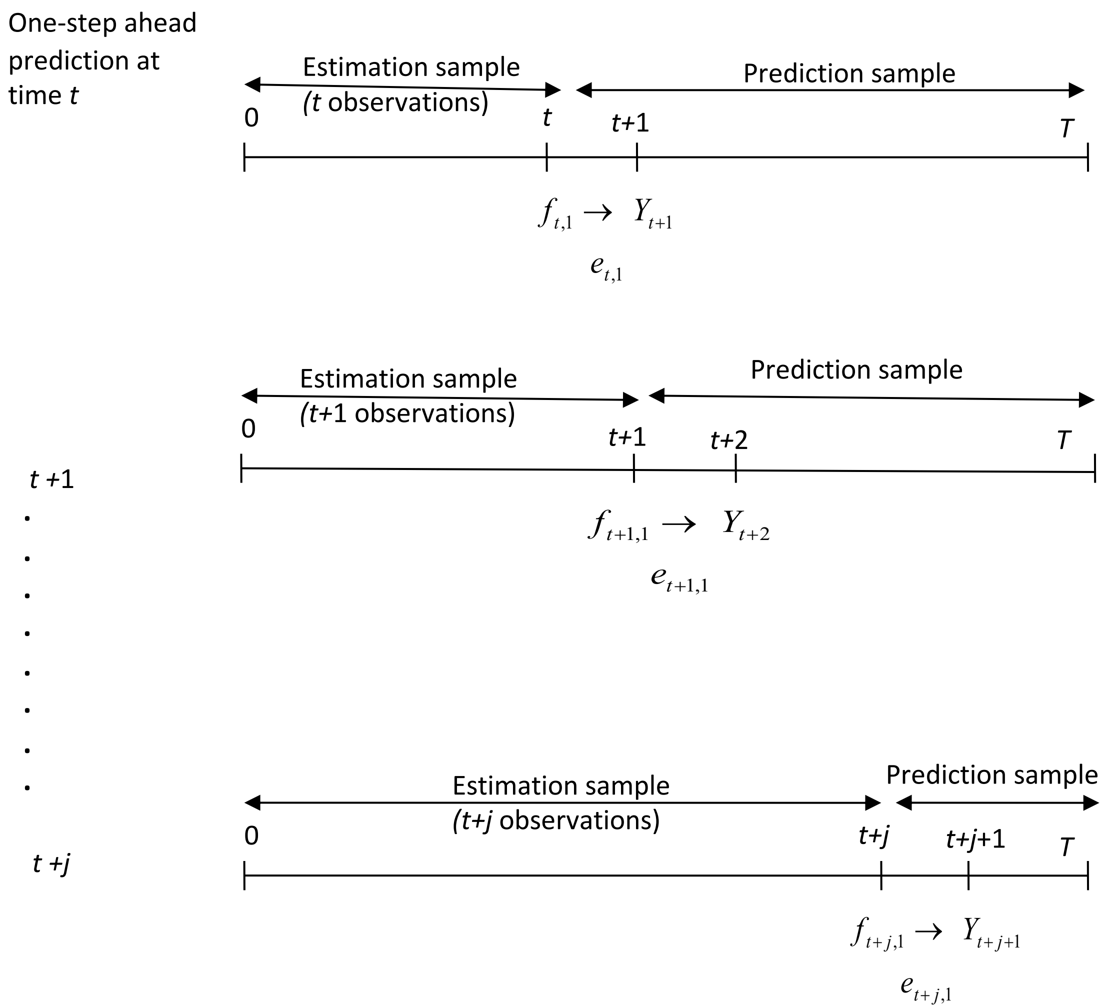
  This sample is used for estimating the model and respective parameters.
- Def: Prediction Sample

  This sample is used to assess the accuracy of the forecast.
- Forecasting Methods:
  - Recursive
  - Rolling
  - Fixed

Forecasting Challenges

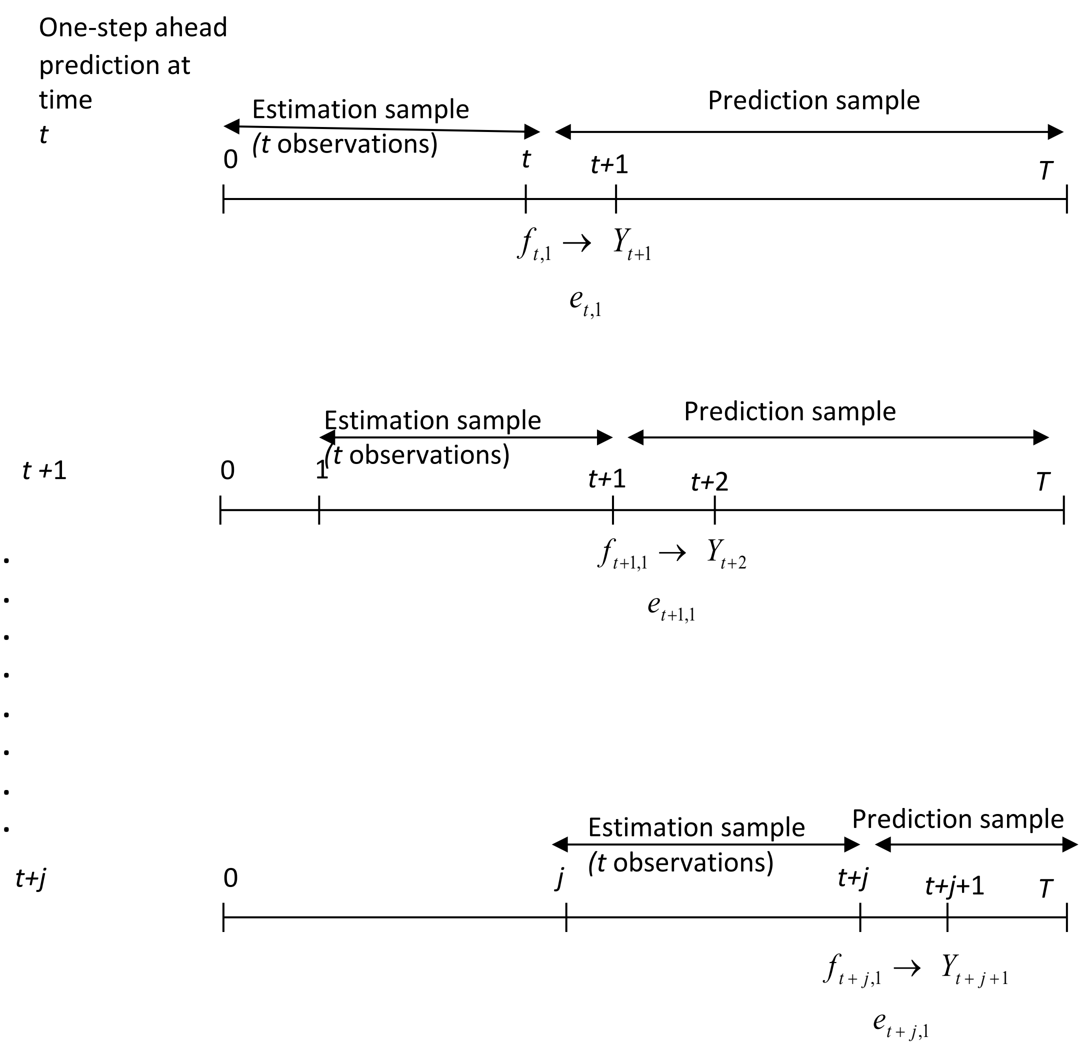
- Lack of understanding of the phenomenon
- Lack of statistical methods
- High uncertainty
- Lack of integration of skills

Recursive Scheme



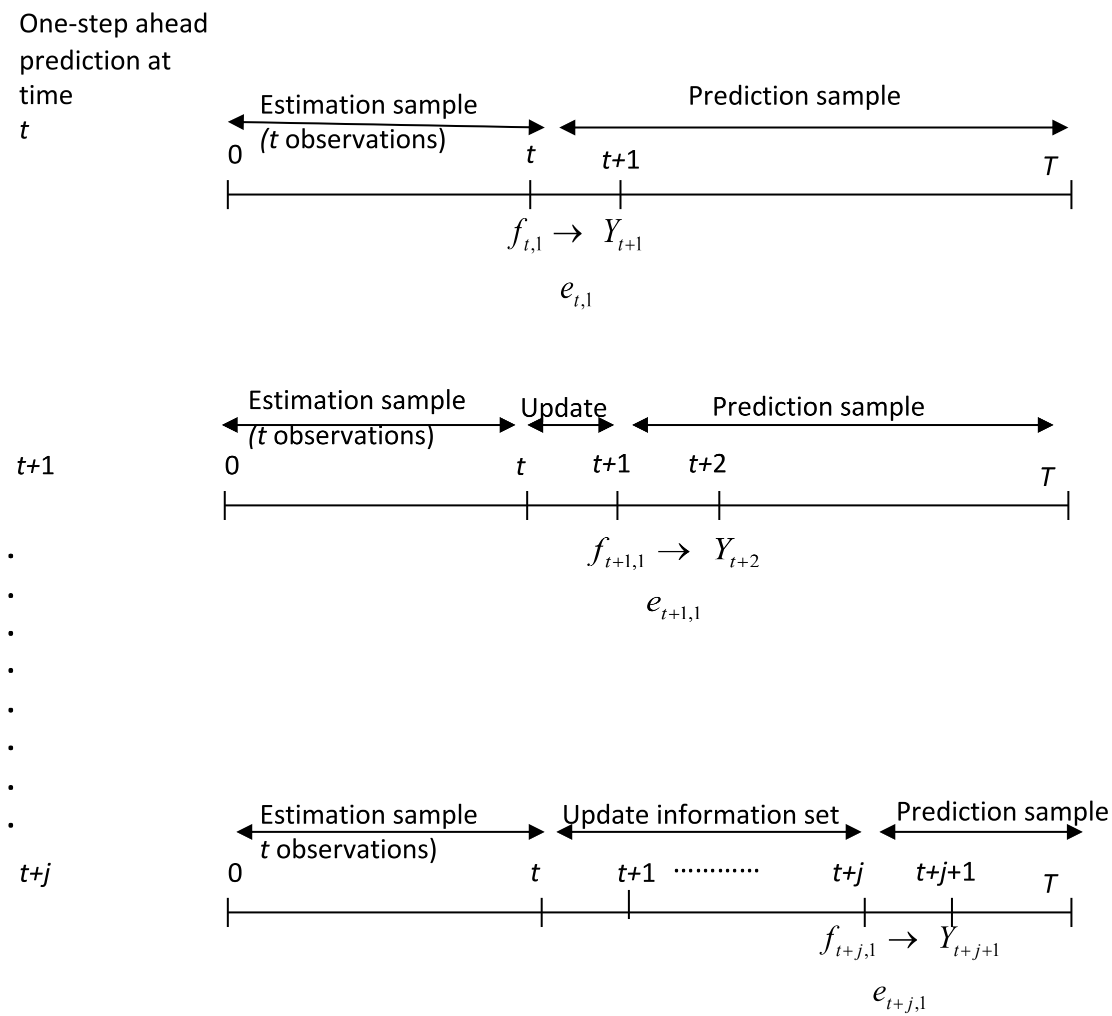
Model parameters are update one observation at a time.

Rolling Scheme



Model parameters are update using a fixed ‘window’ of observations.

Fixed Scheme



Model parameters computed only once

Model Selection

- Among the various model fits, how do we select the best one?
- Need a measure of "best fit model".
- There are many metrics used for model selection such as e.g., MSE, AIC, SIC, Mallows CP, etc.
- Depending on the Forecast problem on hand, certain metrics will be better suited than others for choosing an optimal model.

Model Selection

- Mean Squared Error (MSE):

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

where $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 TIME$ and $e_t = y_t - \hat{y}_t$

- The model with the smallest MSE is also the model with the smallest sum of squared residuals (maximizes R^2).

$$R^2 = 1 - \frac{\frac{1}{T}\sum_{t=1}^{T} e_t^2}{\underbrace{\sum_{t=1}^{T}(y_t - \bar{y})^2}_{\text{Total sum of squares}}}$$

Model Selection

- As the number of parameters increases, the MSE performance deteriorates (overfitting)!

- The out-of-sample forecast will not necessarily improve. However, it will improve the model's fit on the historical data.

- MSE is a biased estimator of the out-of-sample 1-step-ahead prediction error variance.

  → The variance increases as the number of variables increases.

Need to include a penalty for including more degrees of freedom (variables)!

Model Selection

- MSE (adjusted for df): $s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T-k}$

  where k is the number of degrees of freedom (df) used in model fitting.

- Adjusted R^2: $\overline{R} = 1 - \frac{\frac{\sum_{t=1}^{T} e_t^2}{T-k}}{\frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T-1}}$

$$\longrightarrow \overline{R} = 1 - \frac{s^2}{\frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T-1}}$$

Model Selection

- Since: $s^2 = \left(\frac{T}{T-k}\right)\frac{\sum_{t=1}^{T} e_t^2}{T}$

$$\longrightarrow \quad s^2 = \underbrace{\left(\frac{T}{T-k}\right)}_{\text{Penalty Factor}} MSE$$

Model Selection

Two popular model selection metrics are:

$$AIC = e^{\frac{2k}{T}} \frac{\sum_{t=1}^{T} e_t^2}{T}$$ Akaike Information Criterion

$$SIC = T^{\frac{k}{T}} \frac{\sum_{t=1}^{T} e_t^2}{T}$$ Schwarz Information Criterion

Note: SIC is more commonly known as the Bayesian Information Criterion (BIC).

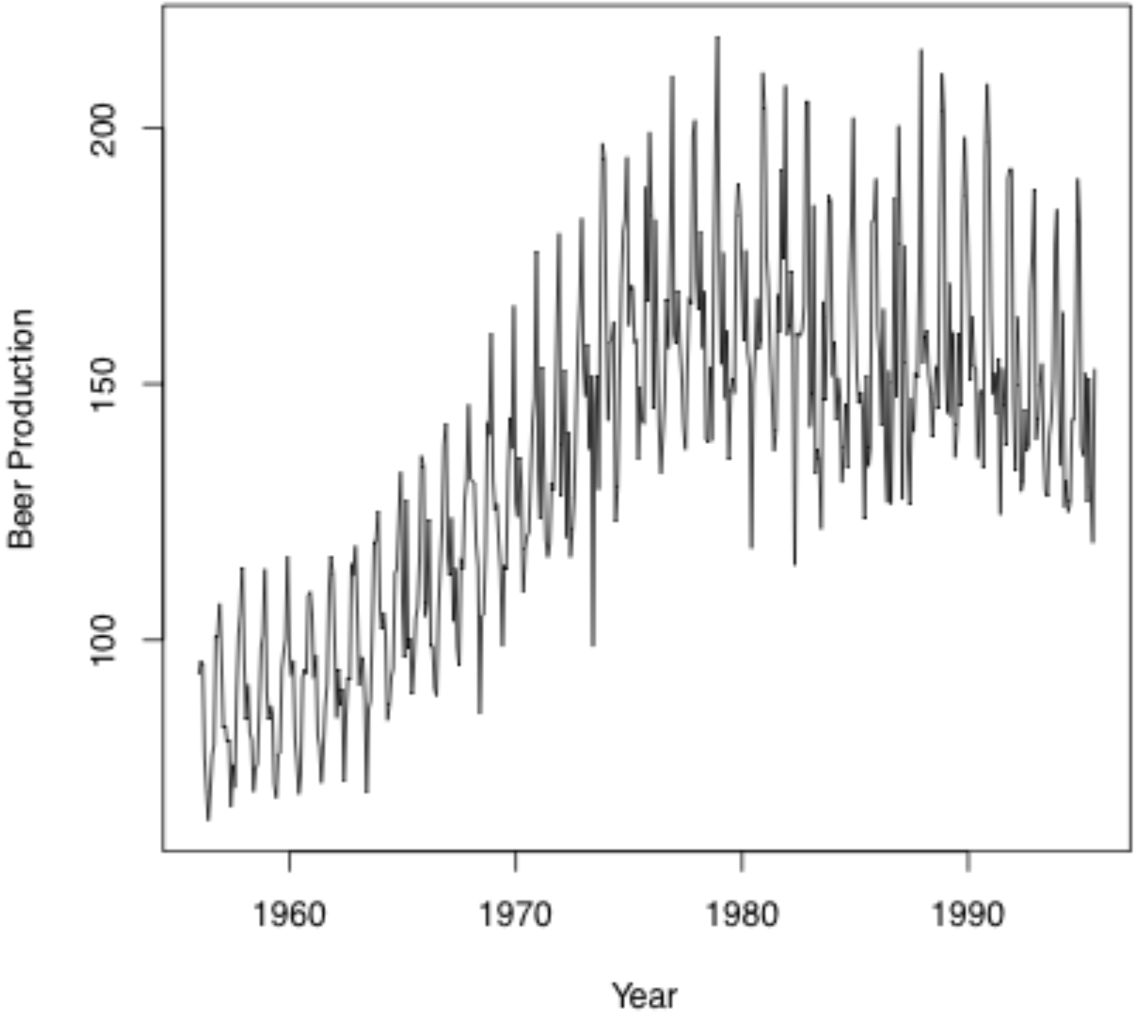
Model Selection

- Consistency: A model selection criterion is consistent if
  1. (a) when the data-generating process (DGP) is among the models considered, the probability of selecting the true DGP approaches 1 as the sample size increases.
  2. (b) when the DGP is *not* among the models considered, the probability of selecting the best approximation to the true DGP, approaches 1 as the sample size increases.
  - MSE: inconsistent
  - AIC: biased towards *overparameterized* models
  - SIC: consistent

- Asymptotic Efficiency: Rate of the model selection process
  - AIC: asymptotically efficient
  - SIC: not asymptotically efficient

Example: Modeling and Forecasting Trend

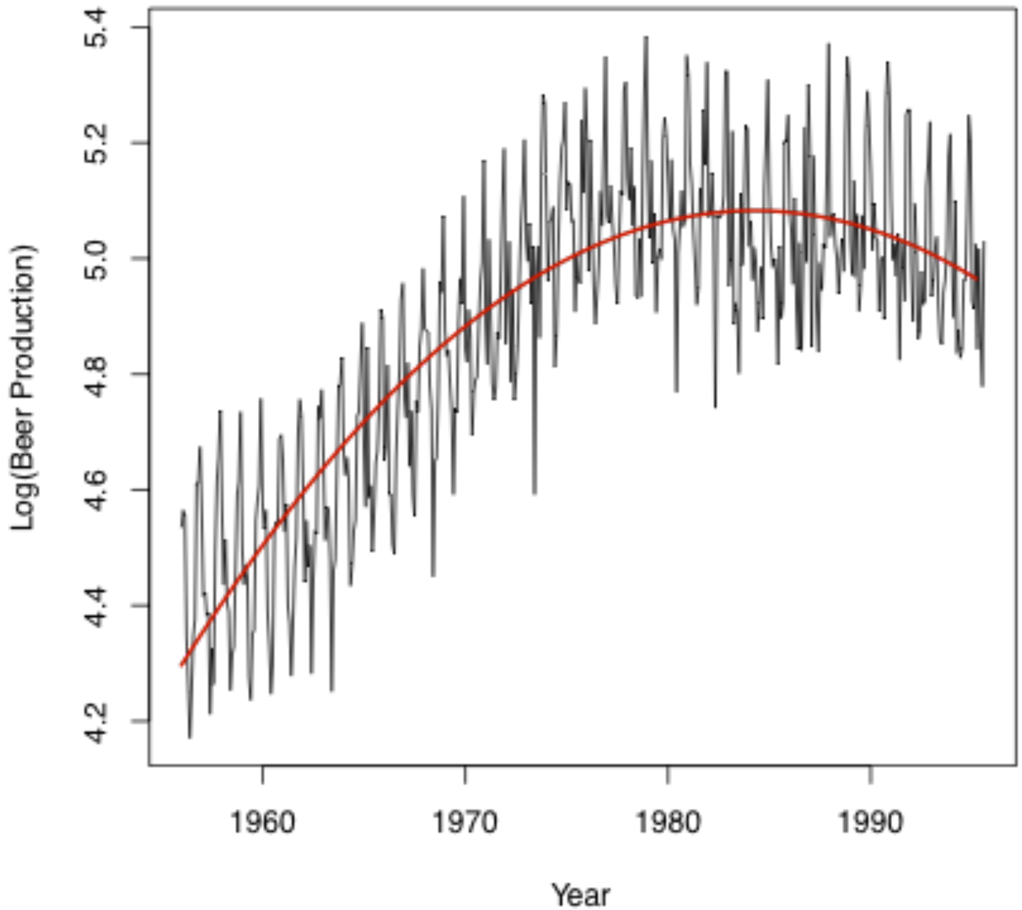
Monthly Beer Production in Australia from Jan 1956 – Aug 1995



http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_intro.pdf

Example: Modeling and Forecasting Trend

Model 1: $\log(y_t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \varepsilon_t$
(Quadratic)



Example: Modeling and Forecasting Trend

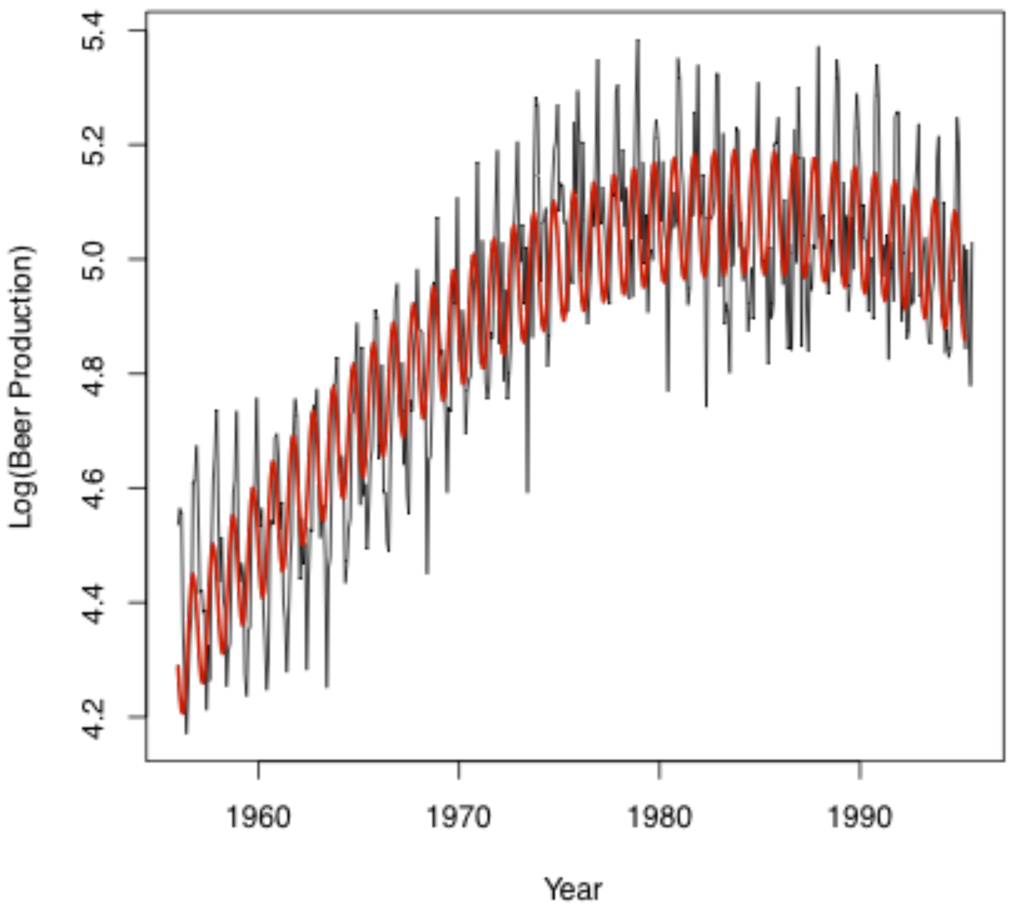
Model 1 (Quadratic): Summary

```
Call:
lm(formula = lbeer ~ t + t2)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.40087 -0.09857 -0.01225  0.09539  0.33826 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.869e+03  2.192e+02  -17.65   <2e-16 ***
t            3.905e+00  2.220e-01   17.59   <2e-16 ***
t2          -9.840e-04  5.618e-05  -17.52   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.141 on 473 degrees of freedom
Multiple R-squared: 0.7176, Adjusted R-squared: 0.7164
F-statistic: 600.9 on 2 and 473 DF,  p-value: < 2.2e-16
```

Example: Modeling and Forecasting Trend

Model 2: $\log(y_t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \boxed{\beta \cos(2\pi t) + \gamma \sin(2\pi t)} + \varepsilon_t$

(Quadratic + Periodic)



Add a periodic term.

Example: Modeling and Forecasting Trend

Model 2 (Quadratic + Periodic): Summary

```
Call:
lm(formula = lbeer ~ t + t2 + sin.t + cos.t)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.33191 -0.08655 -0.00314  0.08177  0.34517 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.833e+03  1.841e+02 -20.815   <2e-16 ***
t            3.868e+00  1.864e-01  20.751   <2e-16 ***
t2          -9.748e-04  4.718e-05 -20.660   <2e-16 ***
sin.t       -1.078e-01  7.679e-03 -14.036   <2e-16 ***
cos.t       -1.246e-02  7.669e-03  -1.624    0.105    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1184 on 471 degrees of freedom
Multiple R-squared: 0.8017, Adjusted R-squared:   0.8 
F-statistic: 476.1 on 4 and 471 DF,  p-value: < 2.2e-16
```

Example: Modeling and Forecasting Trend

Model 1 vs. Model 2

AIC (m1, m2)	df	AIC
Model 1	4	-509.3847
Model 2	6	-673.7203

Model 2 is better.

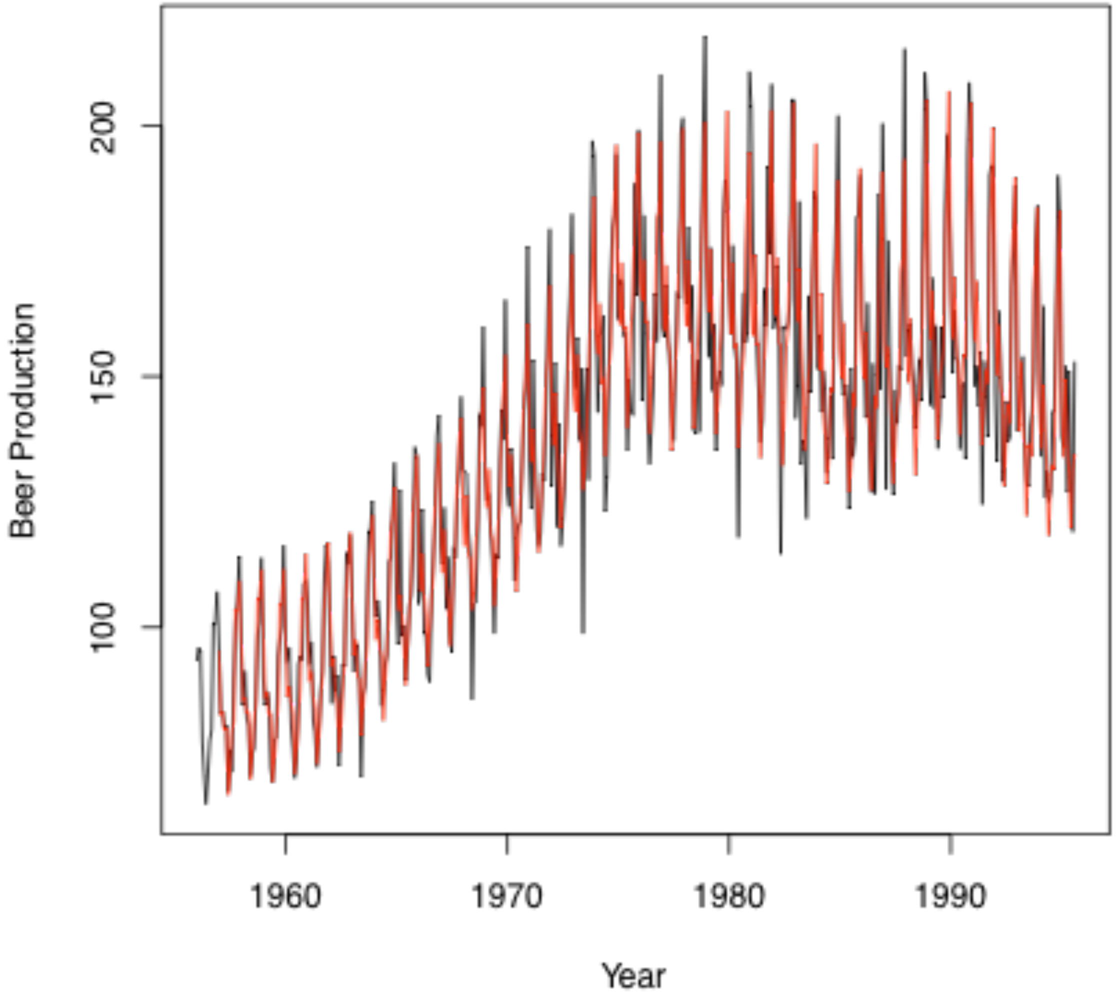
BIC (m1, m2)	df	BIC
Model 1	4	-492.7230
Model 2	6	-648.7278

Model 2 is better.

The smaller the value returned from AIC and BIC, the better the model.

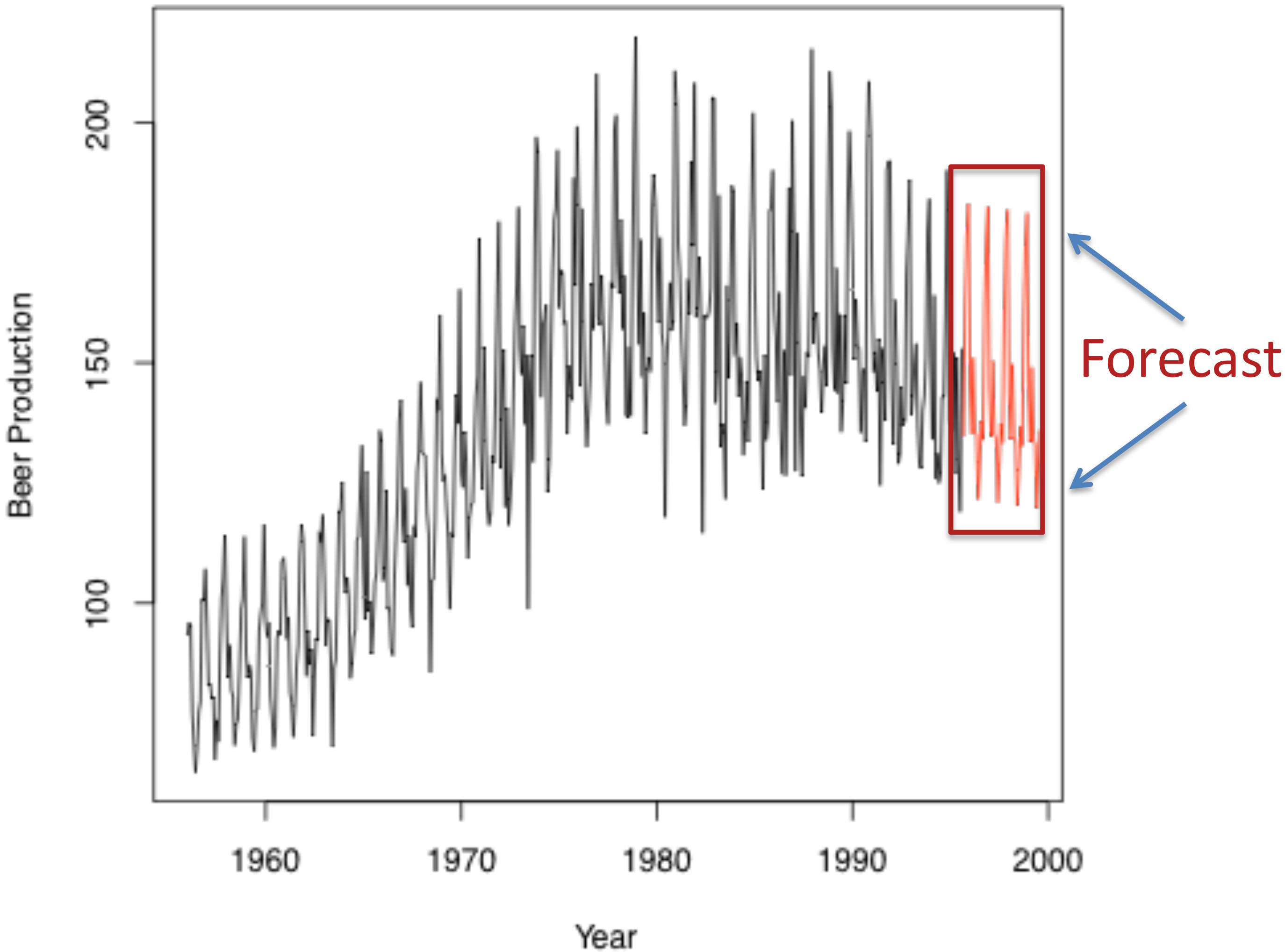
Example: Modeling and Forecasting Trend

Holt-Winters Filter: Considerably better model!



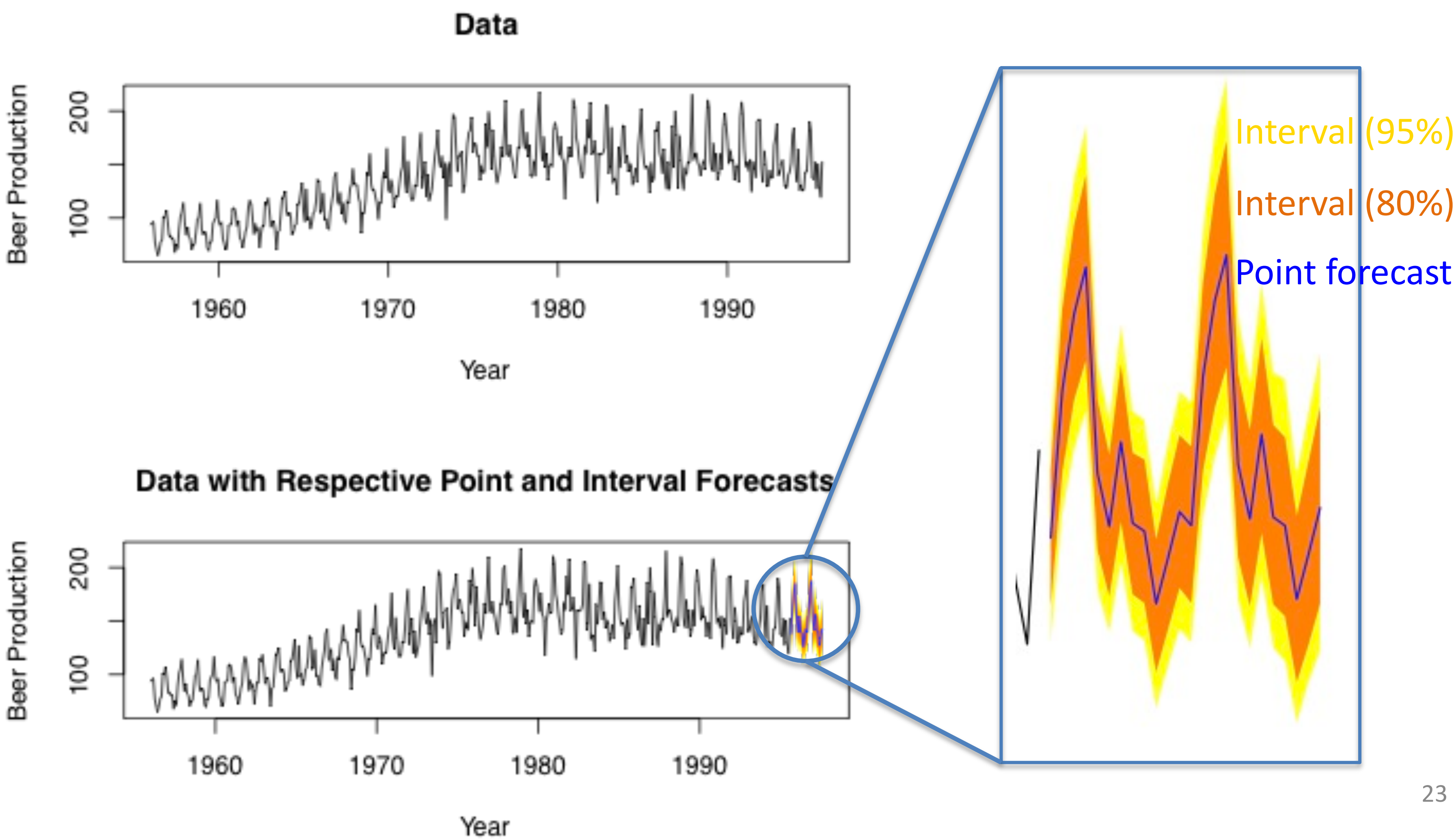
Example: Modeling and Forecasting Trend

Holt-Winters Prediction/Forecast for next 4 years



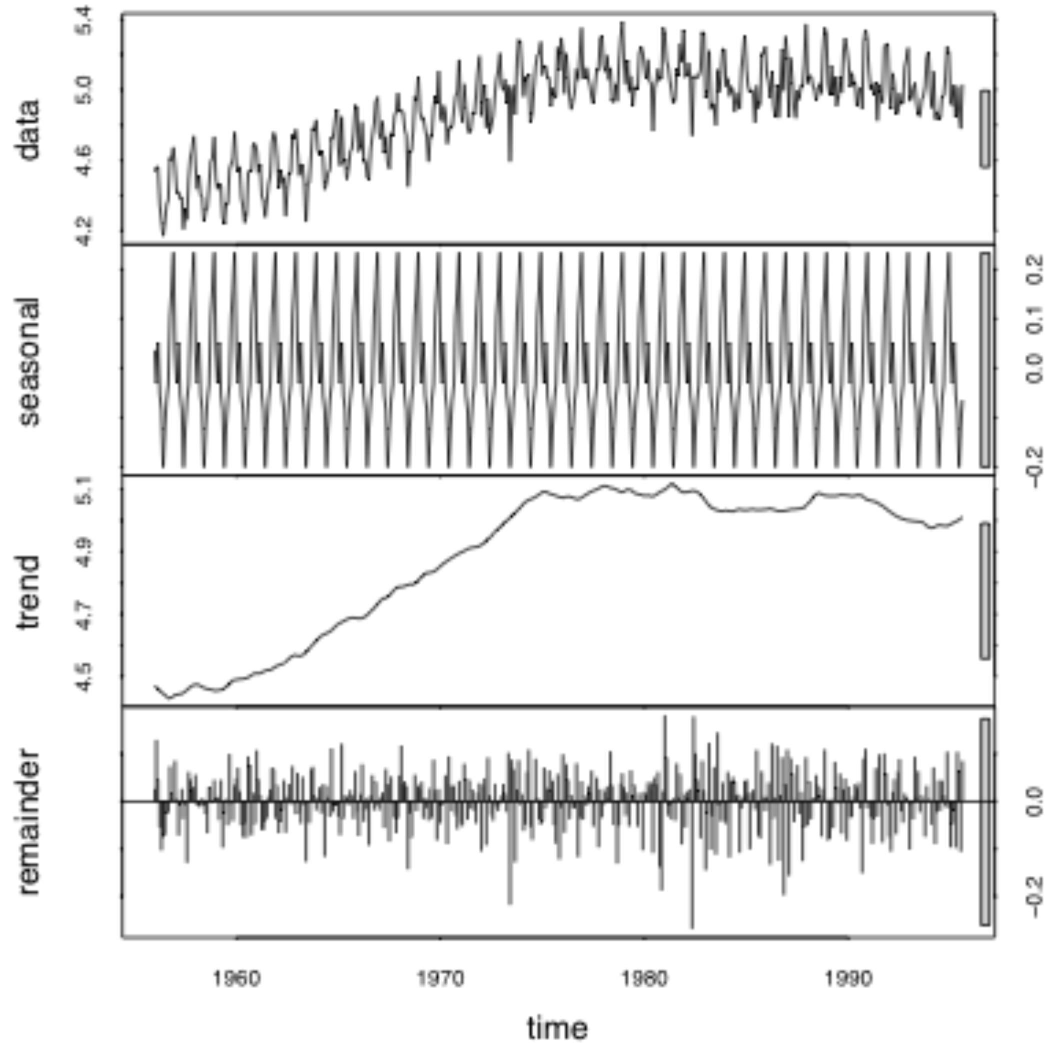
Example: Modeling and Forecasting Trend

Holt-Winters Point and Interval Forecast for next 4 years

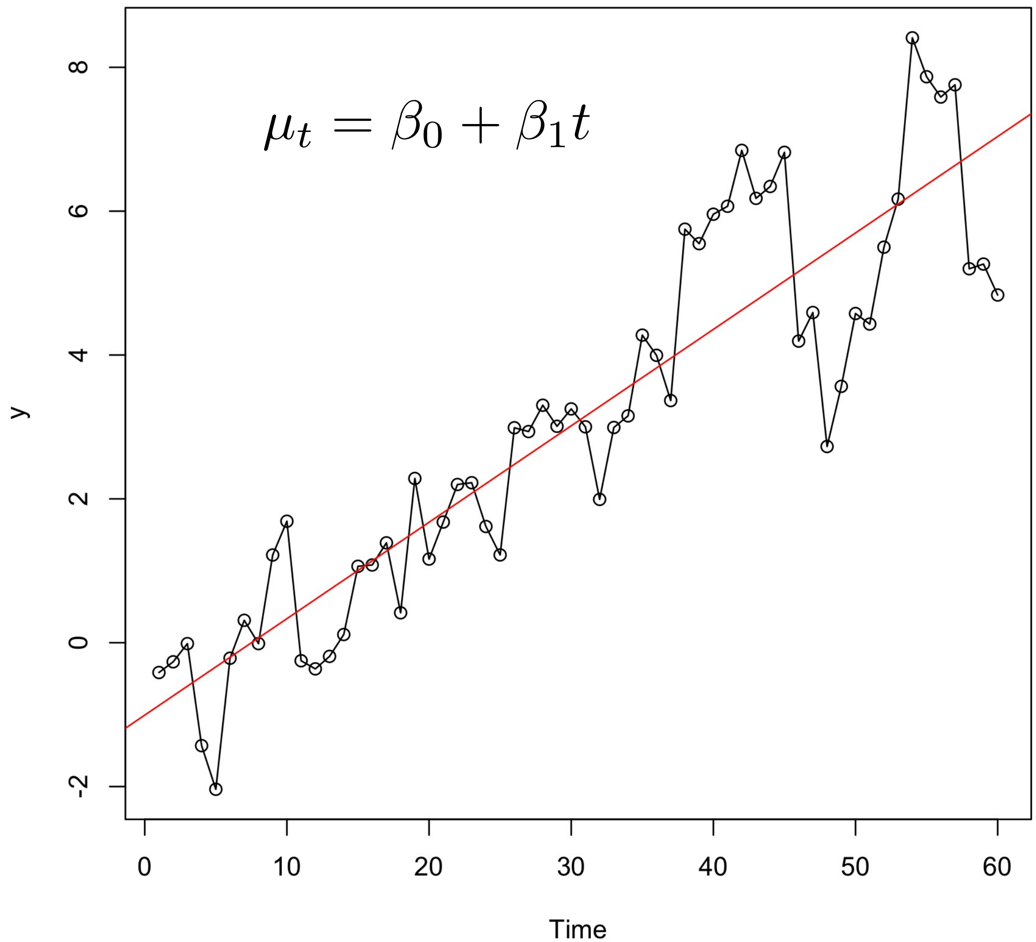


Example: Modeling and Forecasting Trend

Trend + Seasonal Components Decoupled

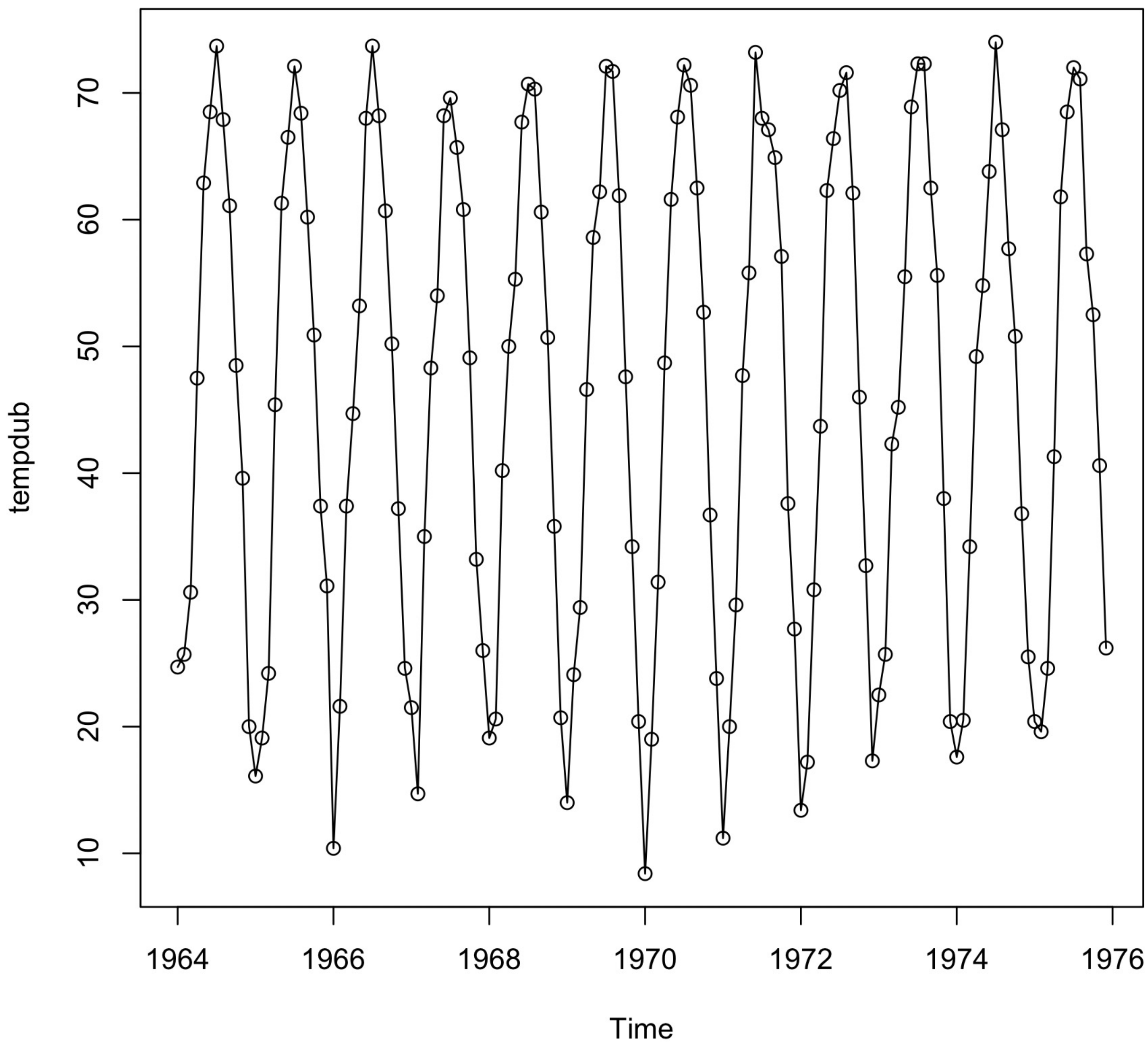


Modeling Trend
Random Walk with Linear Time Trend



	Estimate	Std. Error	*t* value	*Pr*(>|*t*|)
Intercept	−1.008	0.2972	−3.39	0.00126
Time	0.1341	0.00848	15.82	< 0.0001

Modeling Trend
Cyclical or Seasonal Trends



Modeling Trend
Cyclical or Seasonal Trends

$Y_t = \mu_t + X_t$ Represents the series, where $E[X_t] = 0 \; \forall t$

$$\mu_t = \begin{cases} \beta_1 & \text{for } t = 1, 13, 25, \ldots \\ \beta_2 & \text{for } t = 2, 14, 26, \ldots \\ \vdots & \\ \beta_{12} & \text{for } t = 12, 24, 36, \ldots \end{cases}$$

Seasonal Means

Twelve constant parameters giving the expected average temperature for each of the 12 months

Modeling Trend
Cyclical or Seasonal Trends

	Estimate	Std. Error	*t*-value	*Pr*(>|*t*|)
Intercept	16.608	0.987	16.83	< 0.0001
February	4.042	1.396	2.90	0.00443
March	15.867	1.396	11.37	< 0.0001
April	29.917	1.396	21.43	< 0.0001
May	41.483	1.396	29.72	< 0.0001
June	50.892	1.396	36.46	< 0.0001
July	55.108	1.396	39.48	< 0.0001
August	52.725	1.396	37.78	< 0.0001
September	44.417	1.396	31.82	< 0.0001
October	34.367	1.396	24.62	< 0.0001
November	20.042	1.396	14.36	< 0.0001
December	7.033	1.396	5.04	< 0.0001

Modeling Trend
CosineTrends

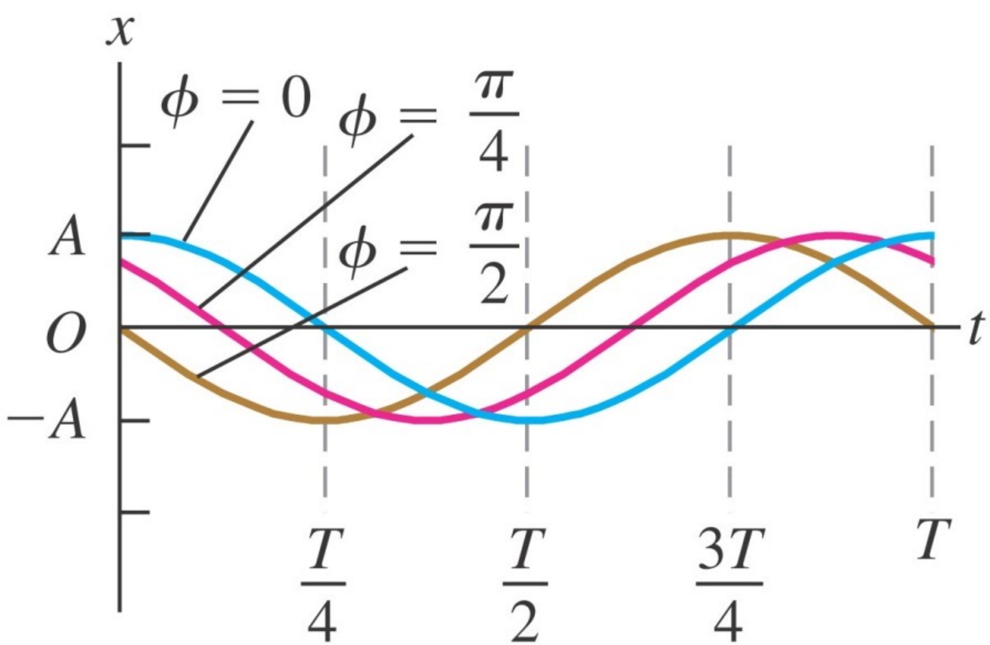
$$\mu_t = \beta\cos(2\pi f t + \Phi)$$

Amplitude (A)

Frequency

Phase

Difficult to estimate because the parameters β, f and Φ are not linear



Modeling Trend
CosineTrends

$$\boxed{\mu_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)}$$

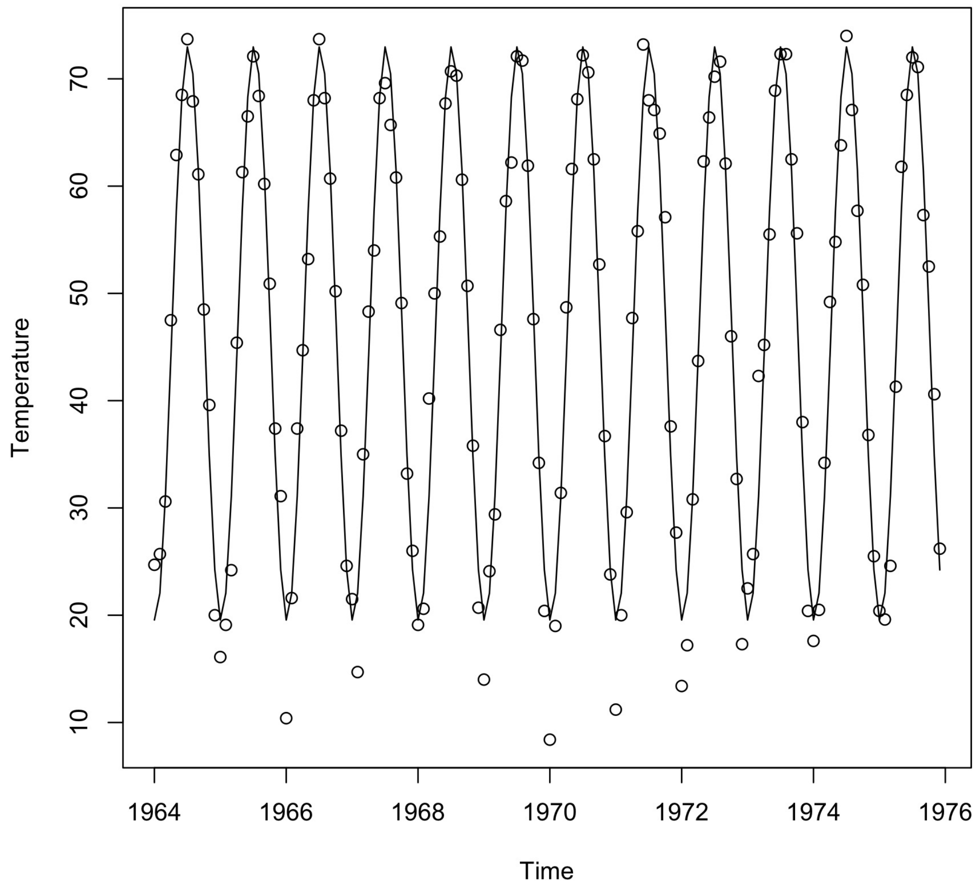
Easier model to estimate

$$\beta \cos(2\pi f t + \Phi) = \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

$$\beta = \sqrt{\beta_1^2 + \beta_2^2}, \qquad \Phi = \operatorname{atan}(-\beta_2/\beta_1)$$

$$\beta_1 = \beta \cos(\Phi), \qquad \beta_2 = \beta \sin(\Phi)$$

Modeling Trend
CosineTrends



Coefficient	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
Intercept	46.2660	0.3088	149.82	< 0.0001
$\cos(2\pi t)$	−26.7079	0.4367	−61.15	< 0.0001
$\sin(2\pi t)$	−2.1697	0.4367	−4.97	<0.0001

Modeling Trend
CosineTrends

EXAMPLE 2.2 Consider the time sequence

$$Z_t = A\sin(\omega t + \theta), \tag{2.1.9}$$

where A is a random variable with a zero mean and a unit variance and θ is a random variable with a uniform distribution on the interval $[-\pi, \pi]$ independent of A. Then

$$\begin{aligned}
E(Z_t) &= E(A)E[\sin(\omega t + \theta)] = 0 \\
E(Z_t Z_{t+k}) &= E\{A^2 \sin(\omega t + \theta)\sin[\omega(t+k) + \theta]\} \\
&= E(A^2)E\left\{\frac{1}{2}[\cos(\omega k) - \cos(\omega(2t+k) + 2\theta)]\right\} \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{2}E\{\cos(\omega(2t+k) + 2\theta)\} \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{2}\int_{-\pi}^{\pi}\cos(\omega(2t+k) + 2\theta)\cdot\frac{1}{2\pi}d\theta \\
&= \frac{1}{2}\cos(\omega k) - \frac{1}{8\pi}[\sin(\omega(2t+k) + 2\theta)]_{-\pi}^{\pi} \\
&= \frac{1}{2}\cos(\omega k),
\end{aligned} \tag{2.1.10}$$

which depends only on the time difference k. Hence, the process is covariance stationary.

Time Series Analysis –W. S. Wei