STA 371G: Statistics and Modeling

Time Series: Fitting a Trend

Mingyuan Zhou
McCombs School of Business
The University of Texas at Austin

http://mingyuanzhou.github.io/STA371G

Time Series Data and Dependence

Time-series data are simply a collection of observations gathered over time. For example, suppose $y_1 \ldots y_T$ are

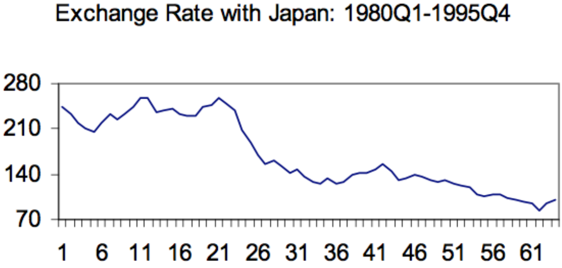
- Annual GDP.
- Quarterly production levels
- Weekly sales.
- Daily temperature.
- 5 minute Stock returns.

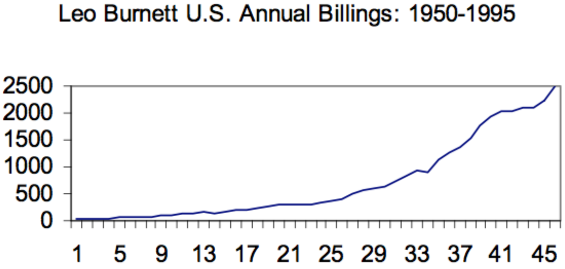
We are usually interested in forecasting with time-series data

Time Series Data and Dependence

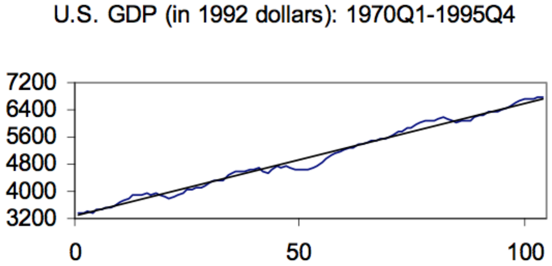
- We might expect what happens at time t to be correlated with what happens at time $t-1$.
- Why do we study time series?
  - Marketing
  - Finance
  - Accounting
  - Transportation
  - ...

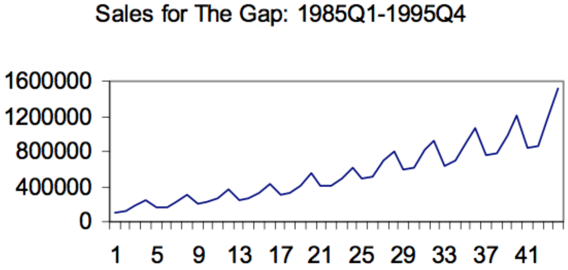
Time Series Data and Dependence





Time Series Data and Dependence



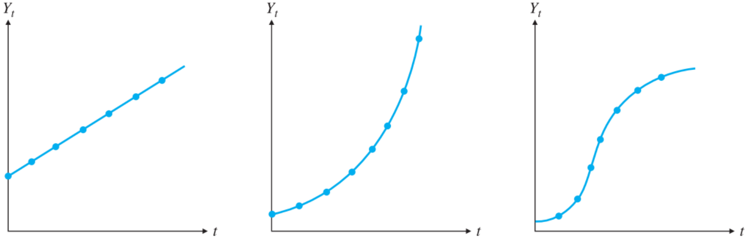


Describe Time Series

Four components of a time series:

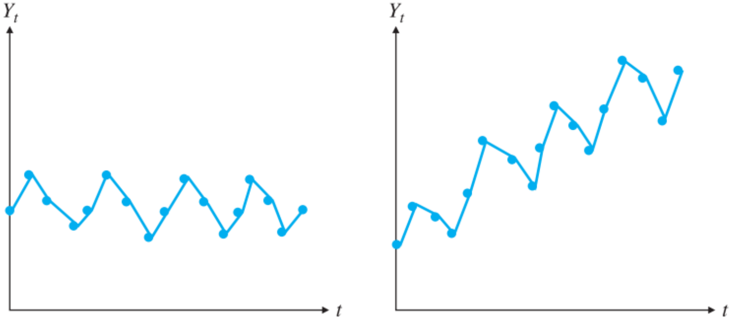
- Trend
- Seasonal
- Cyclic
- Random (noise)

Time Series with Trend



(a) Linear trend (b) Exponential trend (c) S-shaped trend

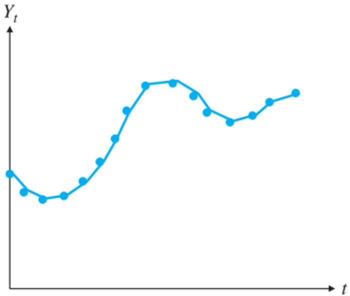
Time Series with Seasonality



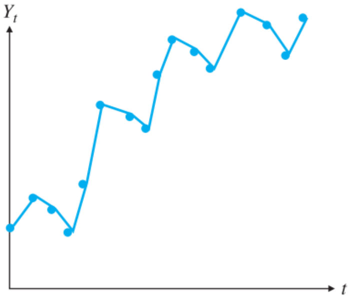
(a) Seasonal component only

(b) Seasonal component with trend

Time Series with Cyclic Component

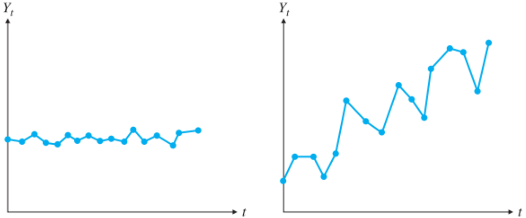


(a) Cyclic component only



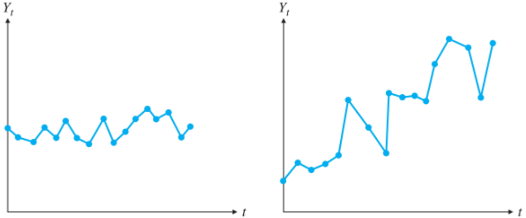
(b) Cyclic component with seasonality and trend

Time Series with Noise



(a) Noise only

(b) Noise superimposed on trend and seasonal components



(a) Noise only

(b) Noise superimposed on trend and seasonal components

Accuracy Measure of Time Series Forecasting

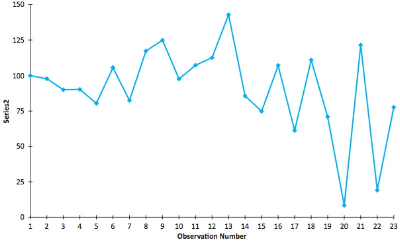
Using $Y_1, \cdots, Y_{t-k}$ to forecast the future value Y_t:

- k-period-ahead forecast: $F_{t-k,t}$
- Forecast error: $E_{t-k,t} = Y_t - F_{t-k,t}$
- Mean absolute error: MAE$=\frac{\sum_{t=1}^{T} |E_{t-k,t}|}{T}$
- Root mean square error: RMSE$=\frac{\sum_{t=1}^{T} E_{t-k,t}^2}{T}$

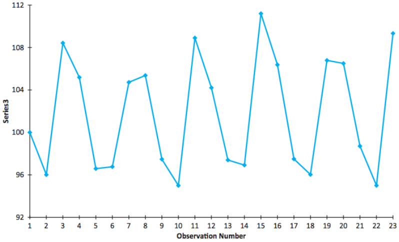
Time series forecasting relies on the assumption that historical patterns repeat themselves. It is likely to fail if this assumption is violated.

Example Time Series

Increasing variance

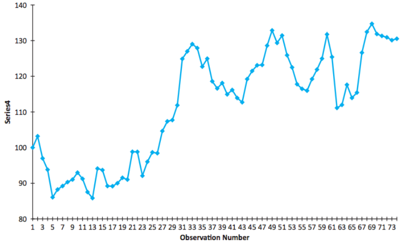


Seasonal

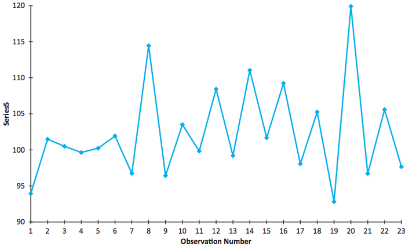


Example Time Series

Meandering



Oscillating frequently



The Run Test (Optional)

Time Series Residuals: 2.2, 9.1, 3.0, -20.1, -2.5, -2 , -1, 4.2, 1.1, 1.2, 1.8, 2, 1

Rune test: $+ + + - - - - + + + + + +$

$N = 13$, $N_+ = 9$, $N_- = 4$, $n = 3$

Under the Hypothesis of Randomness, the number of runs n in a sequence of N elements is approximately distributed as $n \sim \mathcal{N}(\mu, \sigma^2)$, with

$$\mu = \frac{2N_+ N_-}{N_+ + N_-} + 1$$

$$\sigma^2 = \frac{(\mu - 1)(\mu - 2)}{N - 1}$$

If n is significantly larger or smaller than μ, then we may reject the Hypothesis of Randomness.

Time Series Data and Dependence

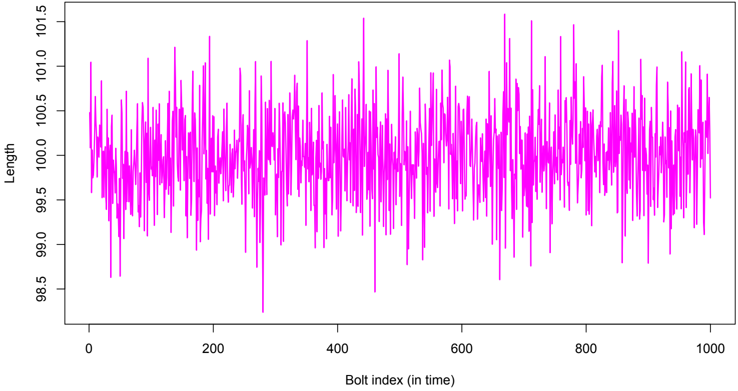
Suppose we measure temperatures daily for several years.

Which would work better as an estimate for today's temp:

- The average of the temperatures from the previous year?
- The temperature on the previous day?

Example: Length of a bolt...

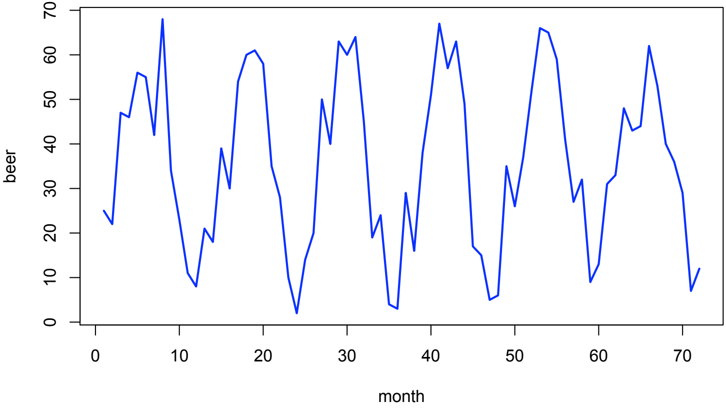
Suppose you have to check the performance of a machine making bolts... in order to do so you want to predict the length of the next bolt produced...



What is your best guess for the next part?

Example: Beer Production

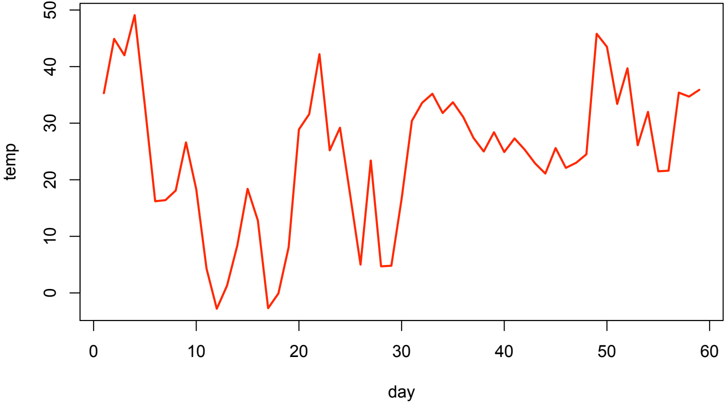
Now, say you want to predict the monthly U.S. beer production (in millions of barrels).



What about now, what is your best guess for the production in the next month?

Examples: Temperatures

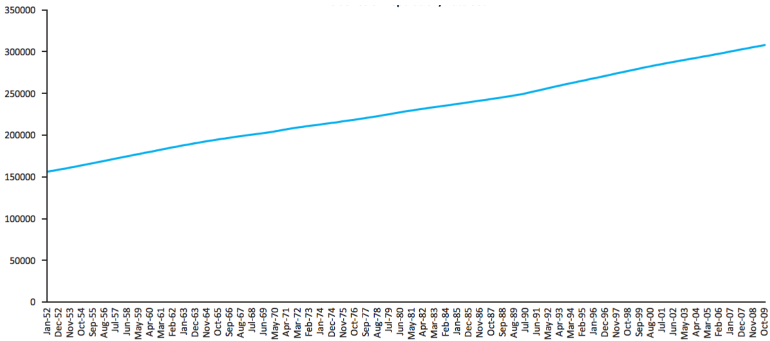
Now you need to predict tomorrow's temperature at O'Hare from (Jan-Feb).



Is this one harder? Our goal in this section is to use regression models to help answer these questions...

Fitting a Trend

Here's a time series plot of US population



What would be a reasonable prediction for US population in October 2015?

Fitting a Trend

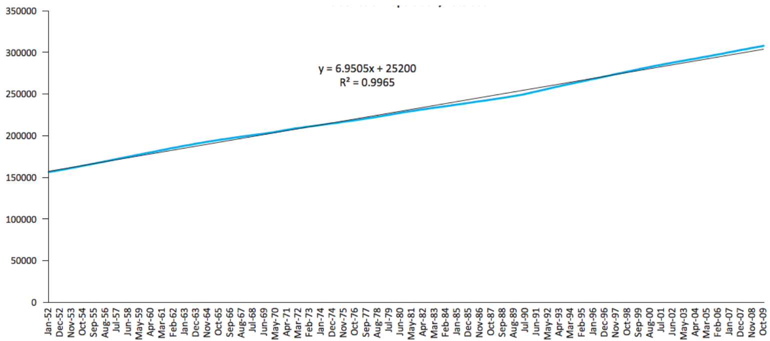
The Population numbers are "trending" upwards... What model could capture this trend?

$$P_t = \beta_0 + \beta_1 t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2)$$

This is a regression of Population (y variable) on "time" (x variable). This allows for shifts in the mean of Population as a function of time.

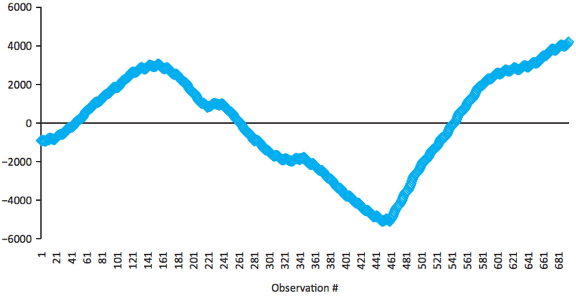
Fitting a Trend

$$P_t = 25200 + 6.9505t + \epsilon_t$$



Fitting a Trend

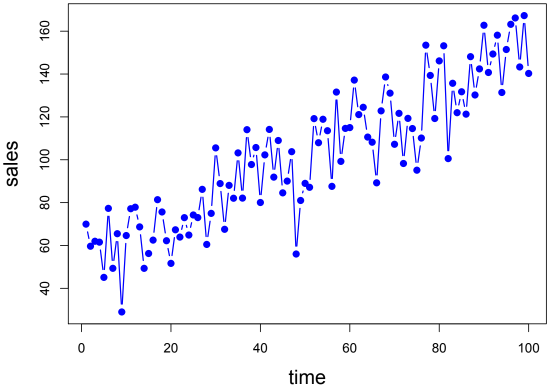
Residuals:



Can we do better?

Fitting a Trend

Here's a time series plot of monthly sales of a company...



What would be a reasonable prediction for Sales 5 months from now?

Fitting a Trend

The sales numbers are "trending" upwards... What model could capture this trend?

$$S_t = \beta_0 + \beta_1 t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2)$$

This is a regression of Sales (y variable) on "time" (x variable). This allows for shifts in the mean of Sales as a function of time.

Fitting a Trend

The data for this regression looks like:

```
months(t)  Sales
1       69.95
2       59.64
3       61.96
4       61.55
5       45.10
6       77.31
7       49.33
8       65.49
...    ...
100     140.27
```

Fitting a Trend

$$S_t = \beta_0 + \beta_1 t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2)$$

Regression Statistics	
Multiple R	0.892
R Square	0.796
Adjusted R Square	0.794
Standard Error	14.737
Observations	100.000

ANOVA

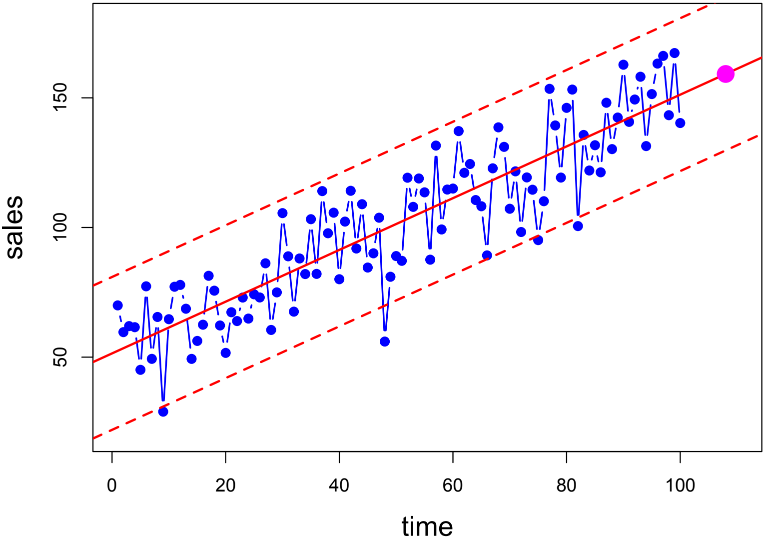
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1.000	82951.076	82951.076	381.944	0.000
Residual	98.000	21283.736	217.181		
Total	99.000	104234.812			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	51.442	2.970	17.323	0.000	45.549	57.335
t	0.998	0.051	19.543	0.000	0.896	1.099

$$S_t = 51.44 + 0.998t \pm 2 * 14.73$$

Fitting a Trend

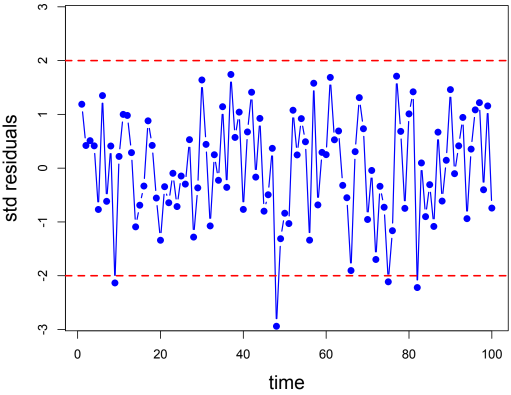
Plug-in prediction...



Residuals

How should our residuals look? If our model is correct, the trend

should have captured the time series structure is sales and what is left, should not be associated with time... i.e., it should be iid normal.



Great!

Exponential Trend

Exponential trend model:

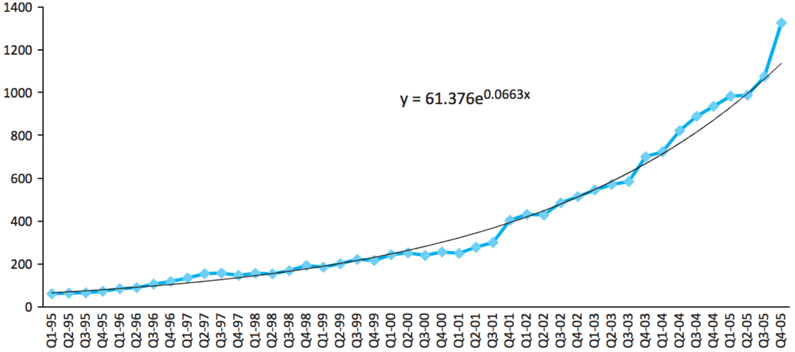
$$Y_t = \gamma_0 \exp(\beta_1 t) u_t$$

Linear trend for $\log(Y)$:

$$\log(Y_t) = \beta_0 + \beta_1 t + \epsilon_t, \ \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Interpretation: Y_t increases by about $100\beta_1$ percent per unit time, as $e^{\beta_1(t+1)} \approx e^{\beta_1 t}(1 + \beta_1)$ if β_1 is close to zero.

Exponential Trend



Exponential Trend