Characteristics of Time Series

- A *time series* is a collection of observations made at different times on a given system.

- For example:
  - Earnings per share of *Johnson and Johnson stock* (quarterly);
  - *Global temperature anomalies* from 1856 – 1997 (annual);
  - Investment returns on the *New York Stock Exchange* (daily).
Digression: Retrieving the Data Using R

jj = scan("http://www.stat.pitt.edu/stoffer/tsa2/data/jj.dat");
jj = ts(jj, frequency = 4, start = c(1960, 1));
plot(jj);

globtemp = scan("http://www.stat.pitt.edu/stoffer/tsa2/data/globtemp.dat");
globtemp = ts(globtemp, start = 1856);
plot(globtemp);

nyse = scan("http://www.stat.pitt.edu/stoffer/tsa2/data/nyse.dat");
nyse = ts(nyse);
plot(nyse);
Correlation

- Time series data are almost always correlated with each other—autocorrelated.

- We may want to exploit that correlation, or merely to cope with it.
**Exploiting Correlation: Forecasting**

- Suppose $Y_t$ is the $t^{th}$ observation, and we observe $Y_0, Y_1, \ldots, Y_{n-1}$. What can we say about $Y_n$?

- If we know the correlation structure, or more precisely the *joint distribution*, of $Y_0, Y_1, \ldots, Y_{n-1}, Y_n$, then we calculate the *conditional* distribution of $Y_n|Y_0, Y_1, \ldots, Y_{n-1}$.

- The conditional mean is the best forecast of $Y_n$, and the conditional standard deviation is the root-mean-square forecast error. If the conditional distribution is normal, we can use them to make probability statements about $Y_n$. 


Coping with Correlation: Regression

- Suppose instead that $Y_t$ is related to a covariate $x_t$, and we are interested in the regression of $Y_t$ on $x_t$.

- Because the $Y$'s are correlated, we should not use Ordinary Least Squares to fit the regression.

- If we knew the correlation structure, we would use Generalized Least Squares.

- Usually we don’t know it, so we must estimate it, typically using a parsimonious parametric model.
Time Domain and Frequency Domain

- Methods that focus on how a time series evolves from one time to the next are called time domain methods.

- Some graphs (e.g. residuals of global temperatures from a quadratic trend) suggest the possibility of waves in the data:

  \[ l = \text{lm}(\text{globtemp} \sim \text{time(globtemp)} + I(\text{time(globtemp)^2})); \]
  \[ \text{plot(globtemp - fitted(l))}; \]

- Since a wave is described in terms of its period, or alternatively its frequency, methods that measure the waves in a time series are called frequency domain methods.