Economic Data Analysis

- Case Studies
  - The *Economist's* Big Mac Price Index
  - Wine Sales in Vancouver BC
  - Kaggle-Walmart Sales Forecasting
  - I Just Run Two Million Regressions (Economic Growth Equation)
  - Variables Selection in a Classification Problem (AmEx Credit Card Default)
  - State Space Time Series Analysis and Forecasting

Big Mac Price Index

- The Big Mac price index was invented by *The Economist* in 1986 as a lighthearted guide to whether currencies are at their "correct" level. It is based on the theory of purchasing-power parity (PPP), the notion that in the long run exchange rates should move towards the rate that would equalize the prices of an identical basket of goods and services (in this case, a burger) in any two countries.

Big Mac Price Index

- Cross Sectional Data
  - Find the latest data for January 2016 over 56 countries
  - Read the data (from Excel spreadsheet)
  - Summarize the data

Big Mac Price Index

- Time Series Data
  - Find the data for US, from 1986 to 2016, unequal-spaced time series
  - Find the data for China, from 1996 to 2016
  - Read the data (from multiple Excel spreadsheets)
  - Summarize and compare

Big Mac Price Index

- Panel Data
  - Find the data for all available countries, from all available years until 2016, unequal-spaced and unbalanced panels
  - Read the data (from multiple Excel spreadsheets)
  - Summarize the data

Wine Sales in Vancouver BC

- Total Weekly Sales of Imported and Domestic Table Wine in Vancouver, BC, Canada from week ending April 4, 2009 to week ending May 28, 2011 (377,228 sales)
  - Data Source: [American Association of Wine Economists](#)

Wine Sales in Vancouver BC

- 377,228 observations of 17 variables in an Excel spreadsheet:

- SKU #, Product Long Name, Store Category Major Name, Store Category Sub Name, Store Category Minor Name, Current Display Price, Bottled Location Code, Bottle Location Desc, Domestic/Import Indicator, VQA Indicator, Product Sweetness Code, Product Sweetness Desc, Alcohol Percent, Julian Week No, Week Ending Date, Total Weekly Selling Unit, Total Weekly Volume Litre

Wine Sales in Vancouver BC

- ## Data Exploration
  - What = Store Category Minor Name (Red/White)
  - Where = Store Category Sub Name (Countries)
  - Price = Current Display Price (Canadian $)
  - Quantity = Total Weekly Selling Unit (Bottles)

- ## Data Visualization
  - Bar, Box, Point, Line, Histogram, Density

- ## Data Analysis
  - Regression: Price Elasticity
  - Classification

Kaggle-Walmart Sales Forecasting

- In early 2014, Kaggle-Walmart announced a recruiting competition using historical markdown data to predict store sales. One challenge of modeling retail data is the need to make decisions based on limited history. If Christmas comes but once a year, so does the chance to see how strategic decisions impacted the bottom line. In this recruiting competition, job-seekers are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains many departments, and participants must project the sales for each department in each store. To add to the challenge, selected holiday markdown events are included in the dataset. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact.

- Since May 2014, the competition has been closed and the winners were announced. For more information about the competition, check here (Note: There are many other interesting on-going competition projects).

Kaggle-Walmart Sales Forecasting

- Data Exploration
  - Cross Section: Store, Department
  - Time Period: Weekly Sales, 2011-2013

- Data Visualization
  - Bar, Box, Point, Line, Histogram, Density

- Data Analysis
  - Regression Analysis
  - Panel Data Analysis

I Just Ran Two Million Regressions

- Following the seminal work of Robert Barro (1991), the recent empirical literature on economic growth has identified a substantial number of variables that are partially correlated with the rate of economic growth.
  - Robert Barro, "Economic Growth in a Cross Section of Countries," QJE (1991), 106(2): 407-443.
  - Xavier X. Sala-i-Martin, "I Just Ran Two Million Regression," AER (1997), 87(2): 178-183.

I Just Ran Two Million Regressions

- Cross Section Data
  - 72 countries, 41 variables
- Economic Growth Equation

$$g = \beta_0 + \beta_y y + \beta_x x + \varepsilon$$

  - g = rates of economic growth
  - y = theoretically important variables
    - Initial level of income; investment rate; school enrollment rate; rate of population growth
  - x = optional or secondary variables

Credit Scoring and Default Prediction

- Based on William Greene, "Sample Selection in Credit-Scoring Models", Japan and the World Economy (1998), 10: 299-316.

- Cross Section Data

    – AmEx: 13,444 obs; 35 variables

- Classification Problem

    – Variable Selection in a Logit Model

$$Y = \Pr(Default \,|\, Card\ Holder) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$$

Chinese Yuan and Stock Market

- High Frequency Time Series
  - Yuan-Dollar Exchange
  - Shanghai Common Stock Market
- Data Visualization
  - Trend, Seasonality
- State Space Time Series Analysis
  - Exponential Smoothing: Trend, Seasonality
  - Time Varying Regression