

Kuan-Pin Lin
Department of Economics
Portland State University

Econometric Computing

- Econometric Computing: Causal Inference of Economic Data for Decision Making
- Econometrics and Predictive Analytics (Machine Learning, Data Mining)
 - Regression and Classification
 - Model Selection
 - Cross-Validation
 - Prediction

Econometric Computing

- Machine Learning from Econometrics
 - Causal Inference Methods
 - Confounding and Instrumental Variables
 - Regression Discontinuity
 - Difference in Difference
- Advanced Topics
 - Non IID Data (Time Series, Panel Data)
 - High Dimensional Sparse Econometric Models

Using R

- Why R (and RStudio)?
 - Powerful, Popular, and Portable
 - Free (Nothing better than that!)
 - But, with steep learning curve!
- Using R Packages
 - `lm()`, `glm()`, `plm()` for linear modeling
 - `tree()`, `rpart()` for machine learning
 - `ggplot2()` for data visualization
 - `parallel()` for multi-core and parallel computing

Cloud Computing with R

- Layers of Cloud Computing
 - Cloud Clients \Leftrightarrow SaaS, PaaS, IaaS
 - IaaS: Infrastructure as a Service
 - PaaS: Platform as a Service
 - SaaS: Software as a Service
- Cloud Services Providers
 - Microsoft Azure
 - Google App Engine
 - Amazon EC2, ...

Econometric Computing in the Cloud

- A Low Cost Option
 - Amazon Web Services (AWS: Free Tier)
 - Elastic Compute Cloud (EC2)
 - Simple Storage Service (S3)
 - [(\$15 + 10c /gb) / month] after 1st year
 - RStudio Server Amazon Machine Image
 - Implementing R / Rstudio AMI
 - [Guide for Dummies](#) (are we?)
 - [Louis Aslett's Guide](#)

An Example

- 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)
 - Irregularly-spaced time series
 - Data Source: [American Association of Wine Economists](#)

An Example

- 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

An Example

- 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

[illegible]

